Computers and Education: Artificial Intelligence 2 (2021) 100018

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.elsevier.com/locate/caeai

Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation

Carlos Felipe Rodríguez-Hernández [a,*], Mariel Musso [b], Eva Kyndt [c,d], Eduardo Cascallar [a]

[a] Center for Professional Learning & Development, Corporate Training and Lifelong Learning, KU Leuven, Belgium, Dekenstraat 2, PB3772, 3000, Leuven, Belgium
[b] Interdisciplinary Center for Research in Mathematical and Experimental Psychology (CIIPME), National Research Council (CONICET), UADE, Argentina, Juan D. Peron, 2158, Buenos Aires, Argentina
[c] Department of Training and Education Sciences, University of Antwerp, Sint-Jacobstraat 2, 2000, Antwerp, Belgium
[d] Centre for the New Workforce, Swinburne University of Technology, Melbourne, Australia

ARTICLE INFO

Keywords:
Artificial neural networks
Prediction
Academic performance
Higher education

ABSTRACT

The applications of artificial intelligence in education have increased in recent years. However, further conceptual and methodological understanding is needed to advance the systematic implementation of these approaches. The first objective of this study is to test a systematic procedure for implementing artificial neural networks to predict academic performance in higher education. The second objective is to analyze the importance of several well-known predictors of academic performance in higher education. The sample included 162,030 students of both genders from private and public universities in Colombia. The findings suggest that it is possible to systematically implement artificial neural networks to classify students' academic performance as either high (accuracy of 82%) or low (accuracy of 71%). Artificial neural networks outperform other machine-learning algorithms in evaluation metrics such as the recall and the F1 score. Furthermore, it is found that prior academic achievement, socio-economic conditions, and high school characteristics are important predictors of students' academic performance in higher education. Finally, this study discusses recommendations for implementing artificial neural networks and several considerations for the analysis of academic performance in higher education.

1. Introduction

Education, similar to many other areas of society and human activity, has been significantly impacted by technological advancements as recent reviews of the last five decades have shown (Chen, Zou, Cheng, & Xie, 2020; Chen, Zou, & Xie, 2020). One such area of sweeping advances has been the application of artificial intelligence in education (AIED). This area has grown over the last 25 years (Roll & Wylie, 2016) and involves newly acquired knowledge from computer sciences and the education field. The simulation of human intelligent behavior using computer implementations has resulted in several educational applications reported in the literature, such as an "intelligent tutor, tutee, learning tool/partner, or policy-making advisor" (Hwang et al., 2020, p. 2). In educational assessment, these new methodologies have improved all aspects of its processes by introducing high-speed data networks and the management of a wide range of data for assessments without the need for traditional testing (Cascallar et al., 2006; Kyndt et al., 2015; Musso & Cascallar, 2009; Musso et al., 2012, 2013; Musso et al., 2020). A relevant research objective in this regard is the evaluation of students' performance and experience (Hwang et al., 2020). Specifically, the prediction of academic performance in higher education provides several benefits to teachers, students, policymakers, and institutions. Achieving such predictive capability with reasonable accuracy could improve the selection process of students who should receive grants or scholarships, could avoid future academic failures and increase retention by providing advanced knowledge on the need for positive interventions, and could serve to identify which teaching practices might have a more positive impact on students' learning (Musso et al., 2013).

Regarding the methodological perspective, Golino and Gomes (2014) report that correlation coefficients, linear and logistic regression analysis, analysis of variance (ANOVA), and structural equation modeling (SEM) are the methods most used to predict students' academic performance. Traditional statistical techniques need to fulfill a set of assumptions such as independence of the observations, homogeneity of the variance,

* Corresponding author.
E-mail addresses: rodriguezcf@gmail.com (C.F. Rodríguez-Hernández), mariel.musso@hotmail.com (M. Musso), eva.kyndt@uantwerpen.be (E. Kyndt), cascallar@msn.com (E. Cascallar).

https://doi.org/10.1016/j.caeai.2021.100018
Received 12 December 2020; Received in revised form 29 March 2021; Accepted 31 March 2021

normality, and linearity (Field, 2009; Tacq, 1997). However, several of these assumptions are not always reported by researchers. Potential undesirable effects of violating statistical assumptions such as type I and type II errors (Nimon, 2012) and inadequate estimation of the effect sizes of statistical parameters (Osborne & Waters, 2002) are ignored. Furthermore, when analyzing large data sets through traditional statistical techniques, the p-values of the predictive models approach zero, so the results are always statistically significant due to the sample size rather than the existence of a truly significant relationship (Khalilzadeh & Tasci, 2017). Therefore, it can be argued that the use of traditional statistical techniques might have introduced some bias when investigating students' academic performance in higher education.

An important contribution from artificial intelligence and the machine-learning area has been the capability to build predictive models of students' academic performance through artificial neural networks (ANNs) (Ahmad & Shahzadi, 2018; Kyndt et al., 2015; Lau et al., 2019; Musso et al., 2012, 2013; Yildiz Yıldız Aybek & Okur, 2018). ANNs entail the possibility of using all the interactions between predictor variables to achieve a better estimation of the outcome variable (Cascallar et al., 2015) and possess the capability of obtaining a prediction even when the independent and dependent variables are related in a nonlinear way (Somers & Casal, 2009). ANNs also allow the analysis of vast volumes of information and the construction of predictive models regardless of the statistical distribution of the data (Garson, 2014). Nevertheless, a lack of conceptual and methodological understanding has prevented an increase in the use of ANNs among educational researchers as they have preferred to fit predictive models based on more traditional approaches such as multiple linear regressions (e.g., Bonsaksen, 2016; De Clercq et al., 2013; Gerken & Volkwein, 2000; Zheng et al., 2002).

In their recent publication, Alyahyan and Düştegör (2020) propose a framework to apply data mining techniques to predict students' academic performance. Alyahyan and Düştegör (2020) state that this framework could promote easier access to data mining techniques for educational researchers, so the advantages of these analysis tools can be fully exploited. As such, three elements should be considered: (1) the definition of academic performance; (2) the predictors of academic performance; and (3) the six stages for building a predictive model, namely, data collection, initial preparation, statistical analysis, data preprocessing, model implementation, and model evaluation.

Based on the framework suggested by Alyahyan and Düştegör (2020), the first objective of this study is to test a systematic procedure for the implementation of ANNs to model two different performance level groups of the SABER PRO test in a cohort of 162,030 Colombian university students. Special focus is given to the *model implementation* and *model evaluation* stages as these stages comprise most of the design decisions that should be made when implementing ANNs. The expected contribution entails the suggestion of several design decisions to implement ANNs so that their use can be further extended among educational researchers.

The second objective of this study is to analyze the relative importance of prior academic achievement, socioeconomic status, high school characteristics, and working status when predicting each performance level group. The expected contribution is to identify the best indicators of future low performers, which could provide useful information as an early warning system in the educational field. Furthermore, understanding which specific patterns of factors lead to future high performance would help to promote these positive conditions.

In line with these general objectives, this study focuses on four specific research questions:

1. What effects are identified during the hyperparameter tuning of ANNs used to classify students' academic performance in higher education?
2. How can error curves be obtained with standard deviation boundaries of ANNs trained to classify students' academic performance in higher education?
3. What is the overall quality of the predictive models of students' academic performance in higher education based on ANNs compared to those based on other predictive methodologies?
4. What is the importance of prior academic achievement, socioeconomic status, high school characteristics, and working status when they are selected as predictors of students' academic performance in higher education?

2. State-of-the-art

2.1. Standardized tests as an indicator of academic performance

Standardized tests can be used in different ways, including accountability, appraisal, and assessment (Morris, 2011). In particular, standardized tests assess the level of knowledge in subject areas (e.g., mathematics, physics, and chemistry) or performance in competencies (e.g., verbal, quantitative reasoning, and writing) (Kuncel & Hezlett, 2007). Tests such as the Scholastic Assessment Test (SAT) administered in the USA (Sackett et al., 2012) or the SweSAT administered in Sweden (Cliffordson, 2008) have been reported to be predictors of further academic performance in universities. As a unique case in higher education, standardized tests are used in Colombia not only to verify students' academic preparation before university but also to measure both the generic and specific competencies they have acquired during their university studies. In this regard, the Colombian Institute for Educational Evaluation (*Instituto Colombiano para la Evaluación de la Educación,* ICFES) designs and manages two national tests for higher education.

The SABER 11 is a standardized test required for admission to higher education institutions in Colombia. The version of the SABER 11 test used in the present study is a cluster of exams that assess students' level of achievement in eight subject areas at the end of high school: biology, physics, chemistry, mathematics, Spanish, English, social sciences, and philosophy. Each subject test consists of multiple-choice questions with only one correct answer and three distractors. The test results are reported using a scale from 0 to 100 for each subject resulting from an item response theory (IRT) analysis of the data. A new version of the SABER 11 test (introduced in 2014) assesses students' performance in the sections of critical reading, mathematics, social sciences, civic competencies, natural sciences, and English; and it maintains the same type of question and scoring system as the previous version of the test (ICFES, 2021a).

The SABER PRO is a standardized test required for graduating from higher education institutions in Colombia. This exam assesses students' performance in the following generic competencies: critical reading, quantitative reasoning, written communication, civic competencies, and English. Each section of the test consists of multiple-choice questions with only one correct answer and three distractors. Besides, the SABER PRO evaluates specific competencies such as scientific thinking for science and engineering students; teaching, training, and evaluation for education students; conflict management and judicial communication for law students; agricultural and livestock production for agricultural sciences students; medical diagnosis and treatment for medicine students; and organizational and financial management for business administration students. The score for each generic competency is reported using a scale ranging from 0 to 300 resulting from an IRT analysis of the data (ICFES, 2021b).

2.2. Predictors of academic performance in higher education

Although an exhaustive review of previous studies on the prediction of academic performance exceeds the goal of this paper, Table 1 summarizes the most recent findings in the field over the last 10 years, highlighting the most important analyzed predictors, the methodological approaches used, and the main reported results (see Table 1). The literature shows a significant increase in machine learning techniques used for academic performance prediction purposes. However, the procedures used in the implementation of these algorithms involve several decision

Table 1
Summary of previous studies on the prediction of academic performance.

Predictors	Method approach	References	Accuracy results
Prior academic achievement (high school results, admission test results, semester GPA, individual course letter marks, and individual assessment grades)	Bayes net, decision tree, k-nearest neighbors, logistic regression, naive Bayes, NN (probabilistic) neural network, rule based, rule induction, random forest, random tree.	Adekitan and Salau (2019); Ahmad et al. (2015); Al-barrak and Al-razgan (2016); Almarabeh (2017); Aluko et al. (2018); Anuradha and Velmurugan (2015); Asif et al. (2017); Asif et al. (2015); Cascallar et al. (2006); Garg (2018); Hamoud et al. (2018); Mesarić and Šebalj (2016); Mohamed and Waguih (2017); Mueen et al. (2016); Musso and Cascallar (2009); Singh and Kaur (2016); Sivasakthi (2017); Yassein et al. (2017)	Predicting at degree level and year level: accuracies of 62% to 89% Predicting success at a course level: accuracies more than 89%. The best accuracy was obtained in course level with 93% (MLP; Sivasakthi, 2017)
Cognitive processes (WM-General Intelligence-Attention-Executive Functions)	Most of them used classical statistical lineal regression. Few studies applied ANNs	Lewandowsky (2011); Kingston and Lyddy (2013); Pérez et al. (2012); Steinmayr et al. (2010); Cascallar et al. (2006); Kyndt et al. (2015); Luft et al. (2013); Musso et al. (2012); Musso et al. (2013); Musso et al. (2020)	A strong relationship between WM and category learning, with a single latent variable. WM mediates category learning across a broad range of tasks. (Lewandowsky, 2011) Sustained attention predicts school performance beyond intelligence (R^2 = .13) Steinmayr et al. (2010) ANN: 83.7% of total accuracy predicting Low Performance (Luft et al., 2013) ANN achieved levels of accuracy of 90–100% (Authors)
Self-regulated Learning Factors (Motivational beliefs, self-efficacy, Learning Strategies, Student interest, behavior of study, stress, anxiety, time of preoccupation)	Most of them used classical statistical lineal regression, Structural Equation Modelling, correlations. Few studies applied machine learning approaches	Bulus (2011); Kingston and Lyddy (2013); Pérez et al. (2012) Garg (2018); Hamoud et al. (2018); Kyndt et al. (2015); Mueen et al. (2016); Musso et al. (2012); Musso et al. (2020);Putpuek et al. (2018)	Academic achievement was predicted by goal orientations and locus of control all together (R^2 = .60) (Bulus, 2011) Self-efficacy and WM predict 19% of the variance in numeracy test scores (Kingston & Lyddy, 2013). The SEM total (R^2 = .659). Ability, effort and learning strategies mediated between academic goals and Academic Achievement (Perez et al., 2012)
Student socio-demographics (Gender, age, race/ethnicity, socioeconomic status)	Bayes net, decision tree, k-nearest neighbors, logistic regression, naive Bayes, NN (probabilistic) neural network, rule based, rule induction, random forest, random tree.	Ahmad et al. (2015); Anuradha and Velmurugan (2015); Garg (2018); Hamoud et al. (2018); Mohamed and Waguih (2017); Mueen et al. (2016); Putpuek et al. (2018); Singh and Kaur (2016); Sivasakthi (2017)	Ahmad et al. (2015): RB-71.3% Anuradha and Velmurugan (2015): NB – 75.2%) Hamoud et al. (2018): REPTree-62.3% Mohamed and Waguih (2017): J48–85.6% Mueen et al. (2016): NB-86% Putpuek et al. (2018): NB - 43.18% Singh and Kaur (2016): J48–67.37% Sivasakthi (2017): MLP-93% Adekitan and Salau (2019): LR - 89.15% Mesarić and Šebalj (2016): J48 – NA
Student's environment (Class type, semester duration, type of program)	Bayes net, decision tree, k-nearest neighbors, logistic regression, naive Bayes (NB), NN (probabilistic) neural network, rule based, rule induction, random forest, random tree.	Adekitan and Salau (2019); Ahmad et al. (2015); Hamoud et al. (2018); Mesarić and Šebalj (2016); Mohamed and Waguih (2017); Mueen et al. (2016)	

Source: Based on Alyahyan and Düştegör (2020) and Musso (2016)

points and the fine-tuning of the model parameters that continue to be applied on a case-by-case basis and are more influenced by individual choice than by consistent methodologies. The present study addresses this methodological gap by systematically exploring the procedure involved in the implementation of ANNs and providing guidelines for the decisions needed at the various stages of data analysis implementation and model evaluation.

On the other hand, a considerable number of studies have been conducted to identify which factors can explain students' academic performance in higher education. These factors have been classified into several theoretical categorizations. For example, McKenzie and Schweitzer (2001) propose four categories: academic, cognitive, demographic, and psychosocial factors. The meta-analysis by Richardson et al. (2012) proposes six categories: demographics (i.e., gender, age, and socioeconomic status), traditional factors (i.e., prior academic achievement), personal traits, motivational factors, self-regulatory learning experiences, and psychosocial contextual influences. An additional classification that encompasses four categories has been proposed by De Clercq et al. (2017) who distinguish past performance, socioeconomic status, self-efficacy beliefs, and study choice. A comparison among these different classifications reveals that prior academic achievement and socioeconomic status are two shared categories of predictors of students' academic performance. However, two additional categories of predictors have also been explored in the literature: high school characteristics (e.g., Black et al., 2015; Win & Miller, 2005) and working status (e.g., Triventi, 2014; Yanbarisova, 2015).

2.3. Artificial neural network (ANN)

An ANN is an information processing system formed by processing units called neurons (Garson, 2014; Haykin, 2009). A neuron, which allows mapping different inputs onto an output, is the most fundamental element in any ANN. To do so, each neuron has a sum block where the inputs are summed after being weighted and an activation function, where the output is calculated after receiving the result from the sum block. The weights represent the strength of the connection between the neurons and the information used by the network to minimize the error with respect to the output (Fausett, 1994). In general, any ANN could be characterized by (1) its arrangement of connections between the neurons (called its topology), (2) its approach to obtaining the strength or the weights on the connections (called its training or learning algorithm), and (3) its activation function (Fausett, 1994).

2.3.1. The multilayer perceptron (MLP)

The ANN topology chosen for the present study is the MLP. MLPs are probably the most common type of ANN and are known for their predictive utility (Garson, 2014). An MLP has a structure consisting of at least three layers: the input layer, the hidden layer, and the output layer. The input layer represents the independent variables or predictors, the hidden layer is where the mapping to relate input and output takes place, and the output layer resembles the dependent variable (Somers & Casal, 2009). MLPs can be used to solve both prediction and classification problems (Garson, 2014; Haykin, 2009). Furthermore, several activation functions can be selected when implementing an MLP. The use of functions such as the hyperbolic tangent or sigmoid allows neural networks to identify complex and nonlinear relationships between the inputs and the outputs (Somers & Casal, 2009).

2.3.2. Training of the MLP

One common way to train an MLP is supervised learning. In classification problems, the purpose of supervised learning is to determine whether the subjects belong to certain groups or categories based on the set of predictors. As such, there is a target category for which the classification accuracy is maximized. As there is also information on both the set of predictors and the corresponding output for each subject, an MLP can learn from these patterns of information. According to Haykin (2009), the supervised learning of an MLP can happen in two different ways: batch and online learning. In *batch learning*, all the cases in the training partition are presented to the MLP at once; while in *online learning*, each case in the training partition is given to the MLP individually. The *online learning* method using supervised learning has been selected for the present study as this method is easy to implement, is capable of tracking small changes in the training data, and is widely used for solving classification problems (Haykin, 2009).

2.3.3. The backpropagation algorithm

The use of the online method with supervised learning for an MLP has increased with the implementation of the backpropagation algorithm, which occurs in two stages (Haykin, 2009). In the forward stage, the predictive weights of the MLP are calculated, and the input signal is transmitted through the layers until reaching the output. Then, in the second stage, an error signal is generated by comparing the output of the MLP with the expected value. This error signal is also propagated layer by layer but in the reverse direction. In this way, an MLP can optimize the value of the previous predictive weights by minimizing the error in each cycle until a certain level of precision is reached. To do so, an MLP uses an optimization function to optimize the weights to reduce the error from the error function. Gradient descent is the optimization function chosen in the present study to minimize the error from the mean squared error function.

Two parameters of an MLP are adjusted during the back-propagation algorithm: the learning rate, which changes the value of the weights in each iteration of the learning process; and the momentum, which increases the speed of the learning process by adding a fraction of the previous weight change to the current weight change (Attoh-Okine, 1999).

2.3.4. Metrics to evaluate the classification obtained by ANN models

There are two different ways to evaluate the quality of the classifications obtained through ANN models. The first is to obtain an error curve during the training phase (Garson, 2014; Haykin, 2009). This can allow one to compare the desired output with the real output provided by the model for each iteration of the training. The second is the use of well-known evaluation metrics based on the confusion matrix of the classification (Saito & Rehmsmeier, 2015). In this respect, rates such as the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates should be obtained to calculate several evaluation metrics (Musso et al., 2013; Saito & Rehmsmeier, 2015). These evaluation metrics are the *accuracy* defined as $(TP + TN)/(TP + FP + FN + TN)$, the *recall* defined as $TP/(TP + FN)$, the *precision* defined as $TP/(TP + FP)$, and the *F1 score* defined as $2(Recall * Precision)/(Recall + Precision)$.

When the classification of interest involves unbalanced data, the *accuracy* value can be misleading, especially when the negative category is the dominant one, since a high *accuracy* value would indicate that only the negative category is being classified correctly (Juba & Le, 2019). That is why it is necessary to calculate additional measures such as the *precision, recall,* and *F1 score*. Since a trade-off exists between the *precision* and *recall* (Alvarez, 2002), it is important to choose which metric provides more information on the quality of the classification provided by ANN models. Such a choice implies penalizing either the FP (a decision on *precision*) or the FN (a decision on *recall*). In this study, we aim to obtain ANNs with high *recall*, leading to classifying as many students as possible who actually belong to each of the performance groups.

3. Material and methods

3.1. Six stages for systematically implementing ANN models

Fig. 1 displays the six stages proposed by Alyahyan and Düştegör (2020), which were followed to answer the first three research questions posed in this study. This section describes each of the stages and explains any design decision made during the implementation of the ANNs.

3.1.1. Data collection

The ICFES provided the data for this study. The original data set contained approximately 200,000 records of Colombian university students who sat for the SABER PRO test in 2016. For each student, there was information about their results from the SABER 11 test, their socioeconomic information, their high school characteristics, and their working status. The original data set also contained additional information such as gender, age, and students' academic program. Note that the students' identity remained anonymous as students were labeled using a code in the data set provided by the ICFES.

3.1.2. Initial preparation

After deleting missing, incorrect, or duplicate data in the original data set, the final sample included 162,030 students of both genders, females (60%) and males (40%), with a mean age of 23.5 years ($SD = 2.89$ years) from both private (64.4%) and public universities (35.6%) in Colombia. The students graduated from academic secondary schools (61.6%), both academic and technical secondary schools (18.1%), technical secondary schools (16.7%), and teacher-training secondary schools (3.6%). Furthermore, the students in the sample were studying one of several different academic programs: agronomy, veterinary medicine, and related degrees (1.3%); arts (3.5%); economics, management, and accounting (25.4%); education (8.7%); engineering, architecture, urban planning, and related degrees (25.6%); health (11.4%); mathematics and natural sciences (2.1%); and social and human sciences (21%). Next, the existing variables in the data set were sorted based on their similarities, leading to the creation of the following categories:

Prior academic achievement. Students' scores in the seven subject areas (biology, physics, chemistry, mathematics, Spanish, social sciences, and philosophy) of the SABER 11 test (between 2006 and 2011) were selected as indicators of prior academic achievement.

Tuition fees. This measure refers to how students pay for their tuition fees. Students pay their tuition fees through either one or several sources of funding: parents, educational loans, own resources, and scholarships.

Students' socioeconomic status. Parents' educational level, parents' occupation, and monthly family income are selected as indicators of students' socioeconomic status (Rodríguez-Hernández et al., 2020; Sirin, 2005; Van Ewijk & Sleegers, 2010).

Students' home characteristics. The number of rooms, whether a student has computer and internet access at home, the type of accommodation (permanent or temporary), and the home stratum are combined under this measure. The home stratum refers to the seven

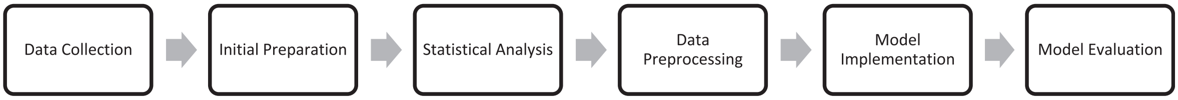

Fig. 1. Stages of the Educational Data Mining framework (Alyahyan & Düştegör, 2020).

categories (ranging from 0 to 6) used by the Colombian government to classify households based on their physical characteristics and the surroundings of the home. The main reason behind this government classification is to hierarchically establish and adjust the price of public services in each area (The World Bank, 2012).

Students' household status. This measure indicates whether a student is the head of a household and, if so, the number of persons the student has in his or her care.

Students' background information. Students' age when taking the SABER PRO test and their gender define their background information.

High school characteristics. Several high school characteristics are part of this measure. These characteristics include the students' high school academic calendar (indicated by the starting month of the academic year), high school "gender composition" (boys-only, girls-only, or coeducational), high school type (public or private), high school schedule (i.e., full-day, morning-only, afternoon-only, evening-only, or weekend-only), and high school diploma (academic, technical, academic-technical or teacher-training).

Working status. The number of hours worked weekly is an indicator of students' working status (Beauchamp et al., 2016; Triventi, 2014).

University background. The number of academic semesters completed by the student before taking the SABER PRO test, university type (e.g., public or private), and students' academic program are analyzed under this measure of university background.

Academic performance in higher education. Students' outcomes in the quantitative section of the SABER PRO test from the 2016 administration were chosen as the operationalization of academic performance in higher education.

3.1.3. Statistical analysis

Appendix A includes the descriptive statistics for the variables analyzed in this study. Both relative and absolute frequencies were calculated for the categorical variables while the means and standard deviations were calculated for the continuous variables. No outliers were found in these analyses. In addition, moderate correlation coefficients were found between the outcomes of the SABER 11 test and the SABER PRO test: Spanish (r(162,030)= .42, $p < .01$), mathematics (r(162,030)= .578, $p < .01$), social sciences (r(162,030)= .451, $p < .01$), philosophy (r(162,030)= .328, $p < .01$), biology (r(162,030)= .469, $p < .01$), chemistry (r(162,030)= .486, $p < .01$), and physics (r(162,030)= .403, $p < .01$).

3.1.4. Data preprocessing

Five steps were followed to preprocess the data set analyzed in this study. In step 1, students' outcomes from the SABER PRO test were categorized using the levels of performance used by the ICFES. In this manner, the categorization of the students' performance levels was based on pre-existing performance standards instead of being classified with a norm-referenced approach. Specifically, the ICFES classifies students' scores from level 1 (lowest performance) to level 4 (highest performance) based on criterion-referenced standards, defining general and specific descriptors of what students are capable of doing at each level of performance, as presented in Table 2. As the level of performance increases, the complexity of the descriptors also increases since a higher level of performance encompasses both the descriptors for that level and the descriptors of all previous levels. As such, we labeled all the students classified in level 1 as the "low performance" group while all the students classified in level 4 were labeled as the "high performance" group. In step 2, students were dummy coded as either belonging or not belonging to each performance group. In step 3, any independent continuous variable was standardized (i.e., converted to a z-score) as they were originally measured using different scales. In step 4, all the categorical variables were dummy coded. Additionally, each continuous variable and each category from the nominal and categorical variables represents an input node for the ANNs. Consequently, there were 122 input nodes (input layer) and 2 output nodes (output layer) in each of the ANNs. In step 5, the total sample was divided into a training set (70%, n = 113,421), a cross-validation set (10%, n = 16,203) and a testing set (20%, n = 32,406).

3.1.5. Model implementation

The software tool chosen for the analyses in this study was NeuroSolutions 7.1, a neural network package that combines a modular, icon-based design interface with the implementation of advanced artificial intelligence and learning algorithms with a user-friendly interface. The *model implementation* stage aimed to systematically train and test ANNs to classify students' academic performance. In this stage, predictive models were developed using the training, cross-validation, and testing sets. The ANNs were trained and tested by following the steps depicted in Fig. 2 while changing the hyperparameters listed in Table 3. The systematic changes to several hyperparameters (i.e., five learning rate values by nine momentum values by three activation functions) led to 135 combinations for training and testing ANNs in each performance group, resulting in 270 combinations. During each training epoch, error curves from both the training and cross-validation sets were generated. When the error curve in the cross-validation set started to increase, it was an indication to stop the training in order to prevent overfitting. In addition, for each combination of hyperparameters, the evaluation metrics described in section 2.3.4 were calculated in the testing set. The selected model in each performance group was the one that possessed the set of hyperparameters that achieved the best performance on the testing set.

3.1.6. Model evaluation

The *model evaluation* stage tested the resulting model for classifying each performance group in two different ways. First, each selected model was trained 30 times again (also for 200 epochs) to generate the error curve with standard deviation boundaries for the training and cross-validation sets. It is important to clarify that the weights of the ANNs were randomly initialized for each new training run. Therefore, the expected subsequent improvement of a previous training was avoided by generating a new training with no prior learning.

Second, several other machine learning algorithms (i.e., fast large margin, decision tree, random forest, and gradient boosted trees) and traditional statistical techniques (i.e., general linear model and logistic regression) were developed using RapidMiner 9.8. Subsequently, the values of the evaluation metrics described in section 2.3.4 were also calculated for these additional predictive methodologies. In this manner, it was possible to compare the overall quality of the classification of students' academic performance provided by the ANNs against the other predictive methodologies.

3.2. The relative importance of the predictors of academic performance in higher education

Sensitivity analysis was conducted to answer the fourth research question posed in this study. Sensitivity analysis has been used in previous research not only as a variable selection method but also to explain a model (Kewley et al., 2000; Yeh & Cheng, 2010). Sensitivity analysis

Table 2
Performance levels for the quantitative reasoning section of the SABER PRO test.

Performance level	General descriptor	Specific descriptors
Level 1 (Score between 0 and 125)	A student at this level identifies explicit information from a single source associated with everyday contexts, which is presented in tables or bar charts containing little data or involving two variables at most.	A student at this level: • Establishes relationships of similarity and order based on the provided information. • Represents in different ways the information contained in tables and graphs.
Level 2 (Score between 126 and 155)	In addition to the above, a student at this level identifies and interprets explicit information from various sources, which is presented in tables and bar graphs, while using simple arithmetic procedures from the given information.	In addition to what was described, a student at this level: • Identifies and extracts explicit information presented in tables and bar graphs. • Represents information contained in bar charts in other record types. • Formulates strategies, validates simple procedures, and solves problems in everyday contexts related to the use of money, business operations, etc., which require the use of: - One or two operations, such as addition, subtraction, or multiplication. - Distributive property of multiplication over addition.
Level 3 (Score between 156 and 200)	In addition to the above, a student at this level extracts implicit information contained in unusual representations associated with the same situation and from a single source of information, argues the validity of procedures, and solves problems using models that combine arithmetic, algebraic, and variational thinking.	In addition to what was described, a student at this level: • Identifies and extracts relevant, explicit, or implicit information presented in unusual graphs, such as stacked bar charts, circular diagrams, etc. • Identifies differences between data representations associated with the same context. • Forecasts results, indicating a single value or a possible interval, based on trends in the presented data. • Proposes strategies and solves problems using percentage calculations, conversion of standard units, simple averages, basic notions of probability or counts that use the principles of addition and multiplication, in a few steps or calculations.
Level 4 (Score between 201 and 300)	In addition to the above, a student at this level identifies and uses implicit information contained in unusual representations from various sources of information to understand a problem situation, argues the validity of procedures, and uses them to solve problems, deciding which one is the most appropriate.	In addition to what was described, a student at this level: • Recognizes the meaning of arithmetic expressions given in the context of solving a problem. • Establishes and uses reference points in the plane, making use of notions of parallelism and rotations. • Proposes representations based on the manipulation and transformation of relevant data in contexts with one or more sources of information. • Proposes strategies and solves problems, in contexts with implicit information, using a conversion of non-standard units, operations with decimals, and the concept of proportionality and the rule of three. • Identifies and corrects errors in proposed procedures as a solution to a problem. • Solves problems that require multiple operations or approximations as part of the solution process. • Validates and compares solution procedures of the same problem and the obtained solutions.

Source: The Colombian Institute for Educational Evaluation (ICFES).

also provides a measure of the relative importance of an input variable by calculating how the output of the ANN changes according to changes in that input while the remaining inputs remain fixed. To begin the sensitivity analysis, every input variable was varied between its minimum value and its maximum value (ranging between 0 and 1 for the dummy coded variables). Subsequently, the relative importance of each input variable was normalized with 1 being the value of the most important variable and the other variables having an importance value measured with respect to the most important one. Next, the input variables were grouped into the categories defined in section 3.1.2, with the importance of each category being calculated by adding the relative importance of all the variables within that category. Finally, a comparison across the two performance level groups was conducted.

4. Results

The results of this study are given in accordance with the research questions. The findings stemming from the *model implementation* and *model evaluation* stages are highlighted, given that the findings of the previous four stages are self-contained in their description presented earlier. This section also includes the results of the analysis of the predictors of students' academic performance in higher education.

4.1. Results from the model implementation stage

The *model implementation* stage aimed to systematically train and test several ANNs to classify students' academic performance. Appendix B includes the results of the training and testing of the ANNs using the 270 combinations of hyperparameters. The different effects found in the *model implementation* stage are described next.

4.1.1. Effects during the training

Three effects were identified during the systematic training of the ANNs. The first effect was related to the training duration. The training duration was the least when the *softmax* was set as the transfer function of the output layer regardless of the values of the learning rate and momentum. This fact could indicate that the *softmax* function converges faster towards the minimum error than the two other transfer functions (*sigmoid* and *linear sigmoid*). However, faster training does not necessarily guarantee lower error values in the error curve.

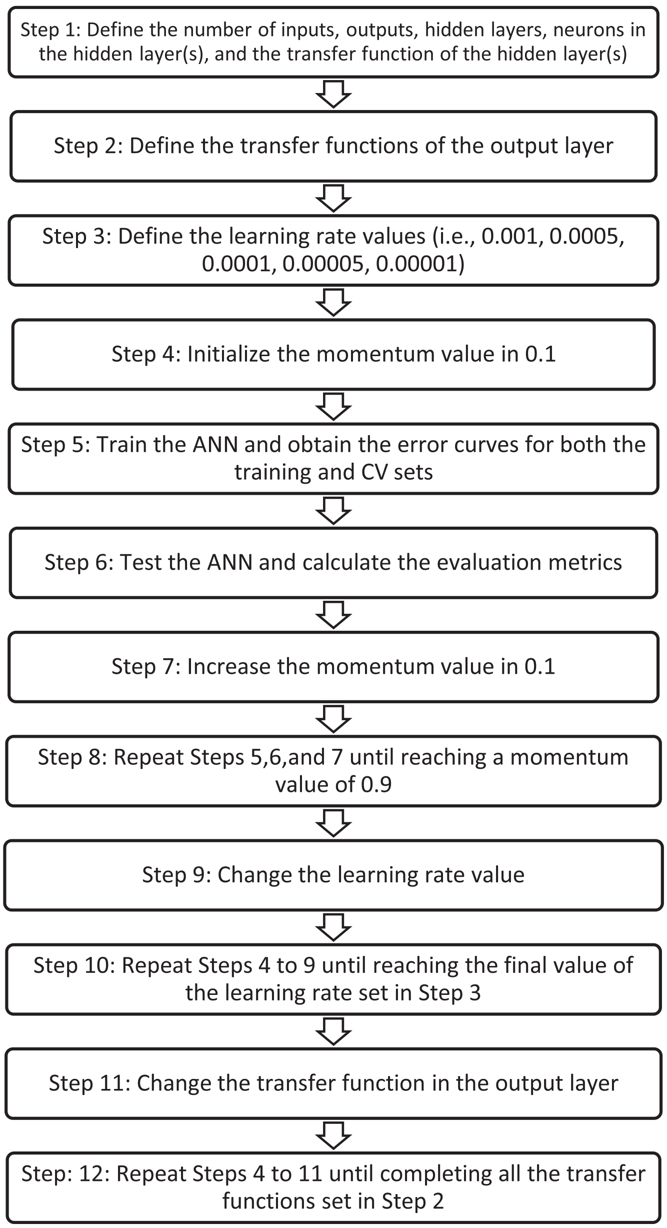

Fig. 2. Tuning of the ANNs hyperparameters.

The second effect was related to the starting value of the error curve. This value was the largest when the *linear sigmoid* was chosen as the transfer function of the output layer with this effect observed in both the training and cross-validation sets. In contrast, when either the *sigmoid* or *softmax* was defined as the transfer function of the output layer, the error curve began from a lower value. This better performance in the error curve might suggest that the *sigmoid* or *softmax* is a more suitable choice for transfer functions in regard to solving classification problems.

The third effect was related to overfitting. When the *linear sigmoid* was chosen as the transfer function of the output layer, the error curve for the cross-validation set increased more rapidly than in the case of the other two transfer functions. Such behavior of the error curve was observed regardless of the learning rate and momentum values. When the *sigmoid* was set as the transfer function of the output layer, the error curves for the training and cross-validation sets were remarkably similar across the 200 epochs, suggesting that there was no overfitting in the model. Altogether, these findings might suggest that a *sigmoid* function converges more effectively towards the minimum of the error curve than the other two transfer functions.

4.1.2. Effects during the testing

The results of the training (Appendix B) revealed that low error values during the training do not necessarily correspond to better evaluation metrics in the testing. An explanation of this finding could be an unbalanced classification due to overtraining. In this context, overtraining can be understood as the lack of discrimination in the predictive system due to the inclusion of random pattern effects during the classification. Therefore, only one category (either "belonging" or "not belonging") had an extremely high value of correctly classified cases.

In the testing, the final model for the "high performance" group correctly classified 82% of the students after setting the values of the hyperparameters as follows: the hyperbolic tangent as the transfer function of the hidden layer, sigmoid as the transfer function of the output layer, a learning rate of .00005, and a momentum of .2. Similarly, the final model for the "low performance" group correctly classified 71% of the students using the following hyperparameter values: the hyperbolic tangent as the transfer function of the hidden layer, sigmoid as the transfer function of the output layer, a learning rate of .0001, and a momentum of .4.

4.2. Results from the model evaluation stage

The *model evaluation* stage assessed the resulting model for classifying each performance group in two different ways. Figs. 3 and 4 display the error curve obtained for each performance group following the procedure described in section 3.1.6. Figs. 3 and 4 illustrate that the error curves for the training and cross-validation sets were quite similar following 30 training runs, also suggesting that there was no overfitting in the models. These findings indicate the consistency in the classification provided by the ANNs. In other words, the results from the classification can be expected to be similar over several training runs of the ANNs.

Table 4 presents the values of the evaluation metrics from the different predictive models for the "high performance" group. The *accuracy* is the lowest in the case of the ANNs. However, higher *accuracy* values could be misleading as there were only 2,260 students in the testing set who belonged to the "high performance" group and 30,146 students who did not (unbalanced data). Although the *precision* is below 60% in all the models, the ANNs detect more false positives, which causes them to achieve the lowest precision value. Conversely, the *recall* and *F1 score* are the highest in the case of the ANNs, resulting in a considerable difference in the recall in favor of the ANNs. These results suggest that the ANNs achieve better performance as they correctly classify most of the students actually belonging to the "high performance" group (higher *recall*) and also achieve a better combined value between the precision and recall (higher *F1 score*).

Table 5 presents the values of the evaluation metrics of the different predictive models for the "low performance" group. The *accuracy* is the lowest in the case of the ANNs. Once again, higher accuracies could be misleading as there were only 6,580 students in the testing set who belonged to the "low performance" group and 25,826 students who did not (unbalanced data). Although the *precision* is below 60% in all the models, the ANNs detect more false positives, which causes them to exhibit the lowest precision value. In contrast, the *recall* and *F1 score* are the highest in the case of the ANNs, again showing a considerable difference in the recall value in favor of the ANNs. These results indicate that the ANNs show better performance as they correctly classify most of the students actually belonging to the "low performance" group (higher *recall*) and also reveal a better combined value between the precision and recall (higher *F1 score*).

4.3. Analysis of the predictors of academic performance in higher education

Table 6 displays the relative and normalized importance of the

Table 3
Hyperparameters of the ANNs.

Topology	Multilayer perceptron (MLP)
Number of input nodes	122
Number of output nodes	2
Number of hidden layers	1
Number of neurons in the hidden layer	50
Number of epochs for training	200
Type of learning	Online
Transfer function of the hidden layer	Hyperbolic tangent
Transfer function of the output layer (x 3)	Linear sigmoid, Sigmoid, and Softmax
Learning rate values (x 5)	.001, .0005, .0001, .00005, .00001
Momentum values (x 9)	Going from .1 to .9 (steps of .1)
Error function	Mean square error
Optimization algorithm	Gradient descent

predictors in the “high performance” group. The outcomes in five sections of the SABER 11 test (mathematics, chemistry, biology, Spanish, and physics), students’ academic program, and monthly family income are the predictors that contribute the most to the classification in the “high performance” group. The predictors were then grouped using the categories presented in section 3.1.2. Fig. 5 provides information on the predictive contribution of each category. The results indicate that prior academic achievement (39.5%), students’ SES (22.8%), university background (15.1%), and high school characteristics (10.2%) are the categories with the largest contribution to the students’ classification in the “high performance” group.

Table 7 displays the relative and normalized importance of the predictors in the “low performance” group. The outcomes in four sections of the SABER 11 test (mathematics, chemistry, biology, and social sciences) and students’ age are the predictors that contribute the most to the classification in the “low performance” group. The predictors for the “low performance” group were also grouped using the categories presented in section 3.1.2. Fig. 6 provides information on the predictive contribution of each category. The findings suggest that prior academic achievement (28.4%), students’ SES (20.3%), university background (10.3%), and high school characteristics (10.2%) are also the categories with the largest contribution to the students’ classification in the “low performance” group.

A comparison across Figs. 5 and 6 reveals the differential contributions of the predictors of academic performance in higher education. Prior academic achievement, students’ SES, and university background provide more information for the classification into the “high performance” group rather than into the “low performance” group. In contrast, students’ home characteristics, how students pay tuition fees, students’ household status, and students’ background have a larger predictive contribution for their classification into the “low performance” group than into the “high performance” group. While high school characteristics are equally important for the classification in both performance groups, working status contributes very little to the classification of either.

5. Discussion

There were two objectives of this study: (1) to test a systematic procedure for implementing ANNs to model two different performance level groups of the SABER PRO test in a cohort of 162,030 Colombian university students and (2) to analyze the relative importance of prior academic achievement, socioeconomic status, high school characteristics, and working status when predicting each performance level group.

Regarding the first research question, the effects found in the model implementation stage can be summarized as follows. First, it is possible to argue that a lower training duration does not imply lower classification error values given that the convergence towards the minimum in the error curve depends on the number of epochs and the transfer function in the output layer. Therefore, when the goal is to maximize the number of correctly classified cases belonging to one specific category, a suitable choice for the transfer function in the output layer is the *sigmoid* function. This is because during training, the *sigmoid* function starts at lower error values and converges more efficiently towards the minimum value of the error curve.

Furthermore, it is not advisable to randomly set the values of the hyperparameters in an ANN model. A random setting is problematic because it might produce undesirable effects in the training of the ANNs such as an increase in the training time or reaching a local minimum in the error curve. A more advisable choice would be to systematically vary the hyperparameters of the ANNs under a controlled training and testing process so that the values for maximizing the classification of interest can be properly reached. As a methodological contribution, the steps proposed in this study for fine-tuning the hyperparameters of ANNs (Fig. 2) can be easily replicated to analyze several other educational data sets.

Regarding the second research question, the present study has demonstrated how to obtain error curves with the standard deviation boundaries of the training of ANNs in a simple manner. Even though there is no rule of thumb for selecting the number of repetitions, the findings of this study suggest that 30 training runs, with 200 epochs each, would provide enough information to graph the error curves for the training and cross-validation sets. These graphs are important as they reveal that the classification provided by predictive systems based on ANNs is stable over several training repetitions. These graphs also make

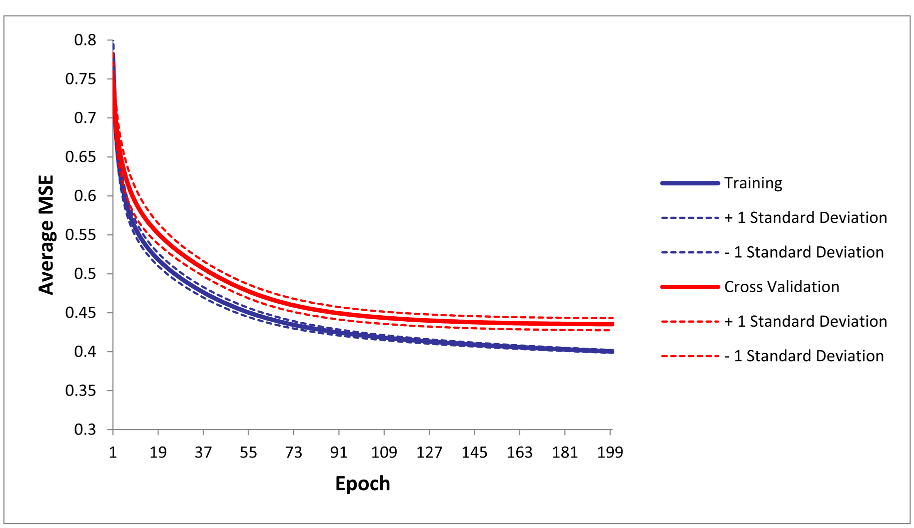

Fig. 3. Error curves for the “high performance” group (lr=.00005, m=.2).

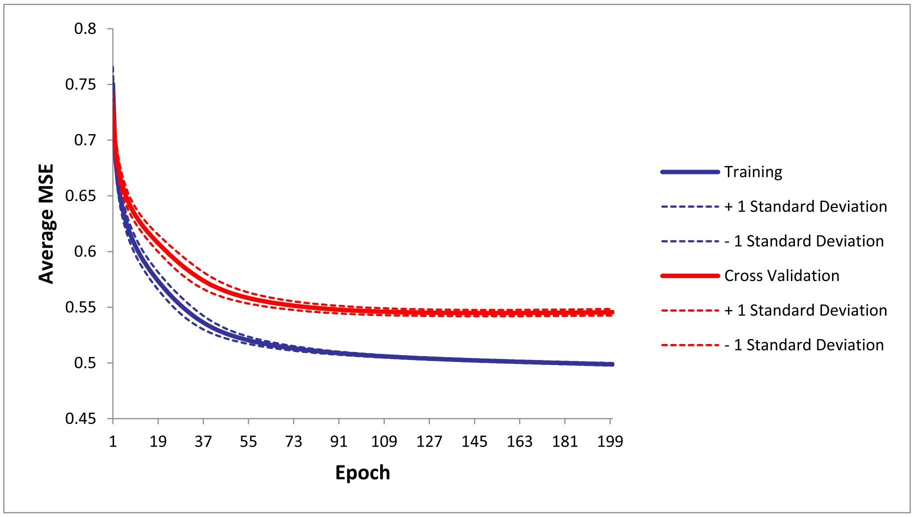

Fig. 4. Error curves for the "low performance" group (lr=.0001, m=.4).

Table 4
Evaluation metrics for the "high performance" group.

	Generalized Linear Model	Logistic Regression	Fast-Large Margin	Decision Tree	Random Forest	Gradient Boosted Trees	ANN
Accuracy	93.5%	93.5%	93.4%	93.0%	93.2%	93.6%	82.1%
Classification error	6.5%	6.6%	6.6%	7.0%	6.8%	6.5%	17.9%
Precision	57.7%	57.5%	60.2%	49.5%	53.7%	60.7%	25.5%
Recall	22.6%	22.9%	16.8%	17.1%	14.7%	18.9%	81.5%
F1 score	0.33	0.33	0.26	0.25	0.23	0.29	0.39

Table 5
Evaluation metrics for the "low performance" group.

	Generalized Linear Model	Logistic Regression	Fast-Large Margin	Decision Tree	Random Forest	Gradient Boosted Trees	ANN
Accuracy	84.2%	84.2%	84.1%	82.4%	82.4%	84.4%	70.8%
Classification error	15.8%	15.8%	15.9%	17.6%	17.6%	15.6%	29.2%
Precision	54.8%	54.8%	54.6%	47.7%	49.3%	58.3%	39.4%
Recall	36.0%	36.0%	35.7%	28.6%	44.7%	33.0%	81.5%
F1 score	0.43	0.43	0.43	0.36	0.47	0.42	0.53

it possible to verify the convergence of ANN models towards a minimum without overfitting.

Regarding the third research question, it is possible to argue that predictive models based on ANNs show good quality in comparison to several other predictive methodologies. In particular, the ANNs implemented in this study exhibited satisfactory values in evaluation metrics such as the *accuracy, recall,* and *F1 score*. The *F1 score* balances the *precision* and *recall*. Therefore, the *F1 score* obtained implies a satisfactory balance between the *precision* and *recall* of the predictive model. Altogether, the results of this study show that predictive models based on ANNs are suitable for solving classification problems with unbalanced data and have advantages over the results obtained when implementing other predictive methodologies.

Regarding the fourth research question, the findings of this study highlight several key points to be considered when predicting students' academic performance in higher education. The first point is that prior academic achievement is the most important predictor of academic performance in universities. While this study corroborates previous evidence on this matter, it was found that prior academic achievement provides more information for the classification of high performers. Belonging to a higher performance group (level 4) involves higher cognitive demands (e.g., to use implicit information from various sources of information to solve a problem) compared with the skills required for low levels of performance (level 1; e.g., to identify explicit information from a single source). If we consider prior academic achievement as a "cognitive proxy" in large-scale assessment (Kuncel et al., 2005; Shaw et al., 2016), this finding is partially consistent with previous studies that have found a greater contribution of basic cognitive variables for low academic performance (Kyndt et al., 2015; Musso et al., 2012, 2013) Nevertheless, prior academic achievement is a top predictor to identify both performance groups.

The second point is that students' SES does contribute to the prediction of their academic performance in universities. This study also suggests the need to consider the level of measurement of SES indicators (i.e., individual, family, or area level) as each level provides a specific contribution to the classification of students' academic performance. The results from this study also indicate that students' SES provides more information to the classification in the "high performance" group than in the "low performance" group. Such a result is consistent with previous research using ANNs (Musso et al., 2012, 2013), which has also found a greater contribution of socioeconomic variables to the identification of high performers.

The third point is that high school characteristics equally contribute to the classification of students' academic performance as being either high or low. Although the predictive weight of high school characteristics could be confounded by other predictors such as prior academic achievement or students' SES, it can be argued that high school characteristics do influence academic performance in universities (e.g., Birch & Miller, 2006; Black et al., 2015; Pike & Saupe, 2002). These results are in agreement with past research using ANNs, which also found a highly

Table 6
Predictive weights for the "high performance" group.

Variable	Predictive weight	Normalized importance
SABER 11 mathematics score	0.1549	100%
SABER 11 chemistry score	0.0741	47.8%
SABER 11 biology score	0.0420	27.2%
SABER 11 Spanish score	0.0411	26.6%
Academic program (Mathematics and natural sciences)	0.0405	26.2%
SABER 11 physics score	0.0402	26.0%
Monthly family income	0.0249	16.1%
SABER 11 social sciences score	0.0247	16.0%
Academic program (Engineering, architecture, urban planning, and related degrees)	0.0230	14.8%
High school academic calendar (B)	0.0201	13.0%
Scholarship (Yes)	0.0189	12.2%
Mother education (Unknown)	0.0186	12.0%
SABER 11 philosophy score	0.0175	11.3%
Father occupation (Unemployed)	0.0144	9.3%
Father education (Graduate level)	0.0139	9.0%
University type (Special regime)	0.0117	7.6%
High school type (Public)	0.0108	6.9%
High school type (Private)	0.0102	6.6%
Mother occupation (Self-employed)	0.0102	6.6%
Father occupation (Unpaid worker for family business)	0.0101	6.5%
Father occupation (N.A.)	0.0097	6.2%
Academic program (Agronomy, veterinary, and related degrees)	0.0091	5.9%
High school academic diploma (Unknown)	0.0089	5.8%
Academic program (Economics, management, and accounting)	0.0089	5.7%
University type (Public municipalities funding)	0.0082	5.3%
Mother occupation (Domestic worker)	0.0082	5.3%
Number of persons in charge	0.0078	5.1%
Mother occupation (Unpaid worker for family business)	0.0077	5.0%
University type (Public departmental funding)	0.0077	5.0%
High school schedule (Evening-only)	0.0075	4.9%
Academic semester	0.0072	4.7%
High school schedule (Full day)	0.0072	4.6%
Academic program (Education)	0.0067	4.3%
Mother occupation (Worker at private business)	0.0065	4.2%
University type (Private corporation funding)	0.0065	4.2%
Mother education (Undergraduate uncompleted)	0.0064	4.1%
Mother occupation (Other activity/ occupation)	0.0062	4.0%
University type (Private foundation funding)	0.0059	3.8%
Mother education (Primary uncompleted)	0.0058	3.7%
Gender (Female)	0.0057	3.7%
Mother occupation (N.A.)	0.0056	3.6%
High school academic diploma (Teacher training)	0.0055	3.6%
Hours worked weekly (Between 11 and 20)	0.0055	3.6%
Resources from parents (No)	0.0055	3.5%
University type (Public national funding)	0.0054	3.5%
Home stratum 6	0.0053	3.4%
Hours worked weekly (Less than 10)	0.0053	3.4%
Mother education (Secondary completed)	0.0053	3.4%
Household of the family (No)	0.0051	3.3%
Mother occupation (Unemployed)	0.0051	3.3%
Home stratum 4	0.0049	3.1%
Mother occupation (Employed at government)	0.0045	2.9%
Educational loan (Yes)	0.0045	2.9%
High school academic diploma (Technical)	0.0045	2.9%
Father occupation (Worker at private business)	0.0044	2.9%
Father occupation (Employer)	0.0044	2.8%
Own resources (Yes)	0.0043	2.8%
High school gender composition (Girls-only)	0.0042	2.7%
Home stratum 0	0.0042	2.7%
High school academic calendar (F)	0.0041	2.7%
Academic program (Social and human sciences)	0.0040	2.6%
Father occupation (Self-employed)	0.0039	2.5%
Own resources (No)	0.0038	2.5%
Resources from parents (Yes)	0.0038	2.5%
Mother education (Graduate level)	0.0038	2.4%
High school gender composition (Boys-only)	0.0037	2.4%
High school schedule (Morning-only)	0.0037	2.4%
Home stratum 2	0.0036	2.3%
Permanent accommodation	0.0035	2.3%
Father education (Undergraduate uncompleted)	0.0035	2.3%
Hours worked weekly (More than 30)	0.0035	2.2%
Mother occupation (Retired)	0.0035	2.2%
Father education (Secondary completed)	0.0034	2.2%
High school academic diploma (Academic and Technical)	0.0033	2.1%
Mother occupation (Laborer)	0.0032	2.0%
Hours worked weekly (Between 21 and 30)	0.0031	2.0%
Mother occupation (Unpaid worker for not family business)	0.0031	2.0%
Hours worked weekly (0)	0.0031	2.0%
Mother education (Technical completed)	0.0030	1.9%
Gender (Male)	0.0028	1.8%
High school schedule (Afternoon-only)	0.0027	1.7%
Academic program (Health)	0.0026	1.7%
Father education (Technical completed)	0.0026	1.7%
PC at home (No)	0.0024	1.5%
Home stratum 3	0.0023	1.5%
Home stratum 5	0.0023	1.5%
Educational loan (No)	0.0022	1.4%
Father education (No education)	0.0022	1.4%
Age	0.0022	1.4%
Father education (Secondary uncompleted)	0.0021	1.3%
Father occupation (Retired)	0.0020	1.3%
High school academic calendar (A)	0.0020	1.3%
Academic program (Arts)	0.0019	1.2%
Father occupation (Domestic worker)	0.0019	1.2%
Temporal accommodation (For studying or any other reason)	0.0019	1.2%
Home stratum 1	0.0018	1.2%
Mother education (Secondary uncompleted)	0.0018	1.1%
Mother education (Undergraduate completed)	0.0017	1.1%
Father occupation (Laborer)	0.0015	1.0%
Father occupation (Unpaid worker for not family business)	0.0015	0.9%
High school schedule (Weekend-only)	0.0014	0.9%
Mother education (Primary completed)	0.0014	0.9%
Father education (Technical uncompleted)	0.0014	0.9%
Mother education (Technical uncompleted)	0.0014	0.9%
Mother occupation (Employer)	0.0013	0.8%
Internet at home (Yes)	0.0013	0.8%
Academic program (Not specified)	0.0013	0.8%
PC at home (Yes)	0.0012	0.8%
Father occupation (Employed at government)	0.0012	0.8%
Father occupation (Other activity/occupation)	0.0011	0.7%
Father education (Primary completed)	0.0010	0.7%
Scholarship (No)	0.0010	0.7%
Father education (Primary uncompleted)	0.0010	0.6%
High school schedule (Ordinary)	0.0010	0.6%
Father education (Unknown)	0.0010	0.6%
High school academic diploma (Academic)	0.0007	0.4%
Father education (Undergraduate completed)	0.0007	0.4%
High school gender composition (Coeducational)	0.0005	0.4%
Mother education (No education)	0.0005	0.3%
Number of rooms at home	0.0005	0.3%
Household of the family (Yes)	0.0005	0.3%
Internet at home (No)	0.0004	0.3%

similar predictive weight of school-related factors to the classification of academic performance as being either high or low (Musso et al., 2020).

The fourth point is that working status does not provide much information for classifying students' academic performance in higher education. However, this result warrants special attention given that working status was merely assessed through a categorical variable

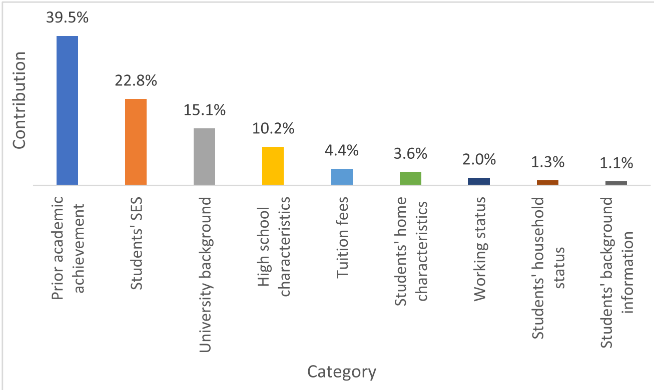

Fig. 5. Categories of predictors in the "high performance" group.

(categories based on the weekly hours worked). In this respect, past educational research (Wang et al., 2010; Yanbarisova, 2015) has suggested defining working status not solely as a dichotomous category but also based on a consideration of additional information on why students work while attending university and other characteristics of their work.

The fifth point is the need to include information on students' academic programs when analyzing their academic performance in universities as students' academic programs provide important information for their classification into different performance groups. A similar finding has been given in the educational literature (e.g., De Clercq et al., 2013; Hansen & Mastekaasa, 2006). A possible explanation for this finding is that different academic programs attract students whose abilities and personal interests are also diverse. Besides, the content of each academic program could have impacted students' readiness and performance level on the SABER PRO test. Therefore, the predictive power of the selected predictors of academic performance might vary across the different disciplines.

5.1. Limitations

The present study supports the use of ANNs in educational research and provides a better understanding of some of the factors involved. However, there are limitations in its scope that must be acknowledged. First, prior academic achievement was found to contribute the most to the classification of students' academic performance. Nevertheless, this predictor was operationalized only using the SABER 11 test outcomes since no other information on students' grades from high school was available. Therefore, relevant information represented via a traditional and widely reported indicator such as high school grade point average (HSGPA) could not be included in this analysis. Furthermore, all of the information regarding students' socioeconomic conditions was self-reported by the students when they were administered the SABER PRO test, a fact that could lead to the inaccurate measurement of such a complex construct as SES. Finally, although the present study used a cross-validation set to avoid the overfitting of the model, more advanced techniques such as regularization and dropout have been reported in the literature on machine learning (e.g., Piotrowski & Napiorkowski, 2013; Srivastava et al., 2014). In particular, the implementation of these techniques in future research could improve the overall quality of the predictive models developed.

5.2. Future research

Future research on the use of ANNs to predict students' academic performance could focus on both educational and methodological topics. Regarding educational topics, indicators of academic performance such as retention or dropout rate could be further explored as outputs of predictive systems based on ANNs. Some initial research (Musso et al.,

Table 7
Predictive weights for the "low performance" group.

Variable	Predictive weight	Normalized importance
SABER 11 mathematics score	0.0694	100%
SABER 11 chemistry score	0.0609	87.6%
SABER 11 biology score	0.0443	63.8%
SABER 11 social sciences score	0.0428	61.7%
Age	0.0392	56.5%
SABER 11 Spanish score	0.0343	49.4%
SABER 11 physics score	0.0208	29.9%
Scholarship (No)	0.0202	29.1%
Number of persons in charge	0.0188	27.1%
Scholarship (Yes)	0.0186	26.8%
Academic program (Engineering, architecture, urban planning, and related degrees)	0.0176	25.4%
Academic program (Mathematics and natural sciences)	0.0168	24.2%
Internet at home (Yes)	0.0163	23.5%
Gender (Female)	0.0162	23.3%
PC at home (Yes)	0.0161	23.2%
High school type (Private)	0.0158	22.7%
High school schedule (Evening-only)	0.0153	22.0%
High school type (Public)	0.0153	22.0%
Internet at home (No)	0.0134	19.3%
Mother education (Unknown)	0.0133	19.2%
Academic program (Education)	0.0128	18.5%
PC at home (No)	0.0118	17.0%
SABER 11 philosophy score	0.0114	16.4%
Hours worked weekly (More than 30)	0.0113	16.3%
High school academic diploma (Unknown)	0.0112	16.2%
Mother occupation (Unpaid worker for family business)	0.0109	15.7%
Educational loan (Yes)	0.0097	13.9%
Monthly family income	0.0092	13.3%
Educational loan (No)	0.0090	13.0%
Mother occupation (Employer)	0.0088	12.7%
University type (Private corporation funding)	0.0083	12.0%
Father occupation (Other activity/occupation)	0.0083	11.9%
Mother occupation (Worker at private business)	0.0082	11.9%
Resources from parents (Yes)	0.0081	11.6%
Father occupation (Unpaid worker for not family business)	0.0080	11.5%
Mother occupation (Laborer)	0.0080	11.5%
Academic program (Social and human sciences)	0.0079	11.4%
Own resources (No)	0.0079	11.3%
Own resources (Yes)	0.0077	11.1%
Academic program (Economics, management, and accounting)	0.0077	11.1%
High school schedule (Ordinary)	0.0076	11.0%
Mother occupation (Self-employed)	0.0075	10.8%
Gender (Male)	0.0072	10.3%
Home stratum 5	0.0071	10.2%
Hours worked weekly (Between 21 and 30)	0.0070	10.1%
Academic program (Arts)	0.0068	9.9%
Mother occupation (Domestic worker)	0.0068	9.8%
Resources from parents (No)	0.0067	9.6%
Mother occupation (Other activity/ occupation)	0.0065	9.4%
Mother education (Undergraduate uncompleted)	0.0061	8.8%
Mother occupation (Unemployed)	0.0061	8.8%
University type (Private foundation funding)	0.0059	8.5%
Number of rooms at home	0.0059	8.4%
Father occupation (Employer)	0.0058	8.3%
Hours worked weekly (0)	0.0057	8.2%
Household of the family (No)	0.0056	8.1%
Academic program (Health)	0.0055	7.9%
Mother occupation (Employed at government)	0.0054	7.7%
Father occupation (Domestic worker)	0.0052	7.5%
Mother occupation (Unpaid worker for not family business)	0.0048	6.9%
Hours worked weekly (Between 11 and 20)	0.0048	6.9%
Father education (Graduate level)	0.0047	6.8%
Father education (Unknown)	0.0046	6.7%

(*continued on next page*)

Table 7 (*continued*)

Variable	Predictive weight	Normalized importance
Academic program (Not specified)	0.0046	6.7%
High school academic diploma (Technical)	0.0045	6.5%
High school academic calendar (A)	0.0045	6.4%
Father education (Undergraduate uncompleted)	0.0044	6.3%
Household of the family (Yes)	0.0044	6.3%
Home stratum 6	0.0043	6.3%
Mother education (Technical completed)	0.0043	6.2%
Home stratum 4	0.0043	6.2%
Father occupation (Retired)	0.0041	5.9%
Temporal accommodation (For studying or any other reason)	0.0041	5.9%
High school academic calendar (B)	0.0040	5.7%
High school schedule (Weekend-only)	0.0040	5.7%
High school academic diploma (Academic and Technical)	0.0039	5.7%
Mother education (Secondary uncompleted)	0.0039	5.6%
Permanent accommodation	0.0038	5.5%
Mother education (No education)	0.0037	5.4%
Father occupation (Laborer)	0.0037	5.3%
Home stratum 3	0.0037	5.3%
Academic program (Agronomy, veterinary, and related degrees)	0.0035	5.0%
Father education (Technical uncompleted)	0.0034	4.9%
Mother education (Technical uncompleted)	0.0034	4.8%
Mother education (Undergraduate completed)	0.0033	4.7%
Home stratum 1	0.0029	4.2%
Father occupation (Self-employed)	0.0029	4.2%
High school academic diploma (Academic)	0.0029	4.1%
Mother education (Secondary completed)	0.0029	4.1%
Mother education (Graduate level)	0.0027	3.9%
Mother occupation (N.A.)	0.0026	3.8%
Home stratum 2	0.0026	3.7%
Father education (Undergraduate completed)	0.0025	3.6%
High school schedule (Full day)	0.0025	3.5%
Father education (Primary completed)	0.0025	3.5%
High school schedule (Afternoon-only)	0.0023	3.4%
High school academic calendar (F)	0.0023	3.3%
Hours worked weekly (Less than 10)	0.0022	3.1%
Mother education (Primary uncompleted)	0.0021	3.0%
Mother occupation (Retired)	0.0019	2.7%
Mother education (Primary completed)	0.0018	2.6%
University type (Public departmental funding)	0.0016	2.2%
High school gender composition (Boys-only)	0.0015	2.2%
Home stratum 0	0.0015	2.1%
Father education (Secondary completed)	0.0015	2.1%
High school gender composition (Girls-only)	0.0014	2.0%
University type (Special regime)	0.0014	2.0%
Father occupation (Employed at government)	0.0013	1.8%
Father occupation (Unemployed)	0.0012	1.8%
University type (Public national funding)	0.0012	1.7%
High school academic diploma (Teacher training)	0.0012	1.7%
Father occupation (Worker at private business)	0.0011	1.6%
University type (Public municipalities funding)	0.0010	1.5%
Father education (No education)	0.0010	1.5%
High school gender composition (Coeducational)	0.0007	1.1%
Father education (Technical completed)	0.0007	1.0%
Father occupation (Unpaid worker for family business)	0.0007	1.0%
Father education (Secondary uncompleted)	0.0007	1.0%
High school schedule (Morning-only)	0.0007	1.0%
Academic semester	0.0005	0.8%
Father occupation (N.A.)	0.0003	0.5%
Father education (Primary uncompleted)	0.0002	0.3%

2020) has demonstrated the value of this methodology in predicting various academic outcomes along a student's academic trajectory. In this way, institutions and policymakers could make targeted efforts to improve the quality of higher education systems and improve several educational outcomes. Furthermore, the complex nature of the relationship between SES and academic performance can be better understood by using nonlinear approaches such as ANNs, as has been shown in this study; and more applications in different populations and circumstances should be explored by adapting the methodology followed in this study to be used for the prediction of academic performance in several other educational settings.

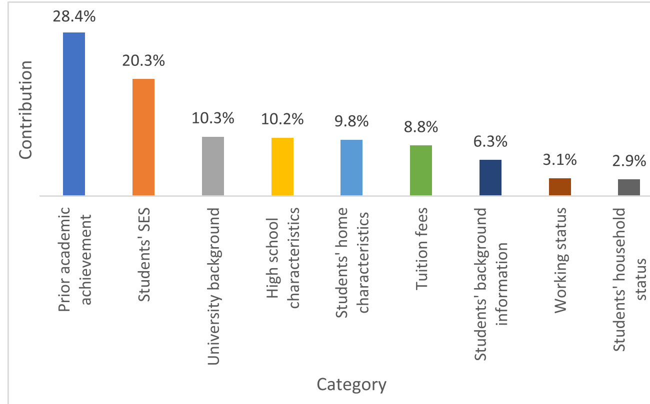

Fig. 6. Categories of predictors in the "low performance" group.

Concerning the methodology used, there is still a clear need for more precise and more generalizable heuristics when using ANNs. In this respect, several parameters of the networks should be studied to better understand their functions in the models. For example, the numbers of hidden layers and neurons in each layer could influence their performance. Furthermore, different ANN topologies could be used to perform tasks such as clustering. Thus, additional topologies based on different learning paradigms should be explored. Finally, the determination of the confidence intervals and the prediction intervals of ANNs also deserves further consideration (e.g., De Veaux et al., 1998; He & Li, 2011).

Statements on open data and ethics

The participants were protected by anonymizing their personal information in this study. The data can be obtained by sending request e-mails to the corresponding author.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Colombian Institute for Educational Evaluation (ICFES) for providing the data for this study. This research was supported in part by the Colombian Department of Science, Technology, and Innovation (COLCIENCIAS) under Grant 756 of 2016.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.caeai.2021.100018.

References

Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon, 5*(2), Article e01250. https://doi.org/10.1016/j.heliyon.2019.e01250

Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematics Sciences, 9*(129), 6415–6426. https://doi.org/10.12988/ams.2015.53289

Ahmad, Z., & Shahzadi, E. (2018). Prediction of students' academic performance using artificial neural network. *Bulletin of Education and Research, 40*(3), 157–164.

Al-barrak, M. A., & Al-razgan, M. (2016). Predicting students' final GPA using decision trees: A case study. *International Journal of Information and Education Technology, 6*(7), 528–533.

Almarabeh, H. (2017). Analysis of students' performance by using different data mining classifiers. *International Journal of Modern Education and Computer Science, 9*(8), 9–15. https://doi.org/10.5815/ijmecs.2017.08.02

Aluko, R. O., Daniel, E. I., Oshodi, O. S., Aigbavboa, C. O., & Abisuga, A. O. (2018). Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology, 16*(3), 385–397. https://doi.org/10.1108/jedt-08-2017-0081

Alvarez, S. A. (2002). An exact analytical relation among recall, precision, and classification accuracy in information retrieval (Technical Report BCCS-02-01). http://www.cs.bc.edu/~alvarez/APR/aprformula.pdf.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in HigherEducation, 17*, 1–21. https://doi.org/10.1186/s41239-020-0177-7

Anuradha, C., & Velmurugan, T. (2015). A comparative analysis on the evaluation of classification algorithms in the prediction of students' performance. *Indian Journal of Science and Technology, 8*(15), 1–12. https://doi.org/10.17485/ijst/2015/v8i15/74555

Asif, R., Merceron, A., Abbas, S., & Ghani, N. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education, 113*, 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

Asif, R., Merceron, A., & Pathan, M. K. (2015). Predicting student academic performance at degree level: A case study. *International Journal of Intelligent Systems and Applications, 7*(1), 49–61. https://doi.org/10.5815/ijisa.2015.01.05

Attoh-Okine, N. O. (1999). Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Advances in Engineering Software, 30*(4), 291–302. https://doi.org/10.1016/S0965-9978(98)00071-4

Beauchamp, G., Martineau, M., & Gagnon, A. (2016). Examining the link between adult attachment style, employment, and academic achievement in first semester higher education. *Social Psychology of Education: International Journal, 19*(2), 367–384. https://doi.org/10.1007/s11218-015-9329-3

Birch, E. R., & Miller, P. W. (2006). Student outcomes at university in Australia: A quantile regression approach. *Australian Economic Papers, 45*(1), 1–17. https://doi.org/10.1111/j.1467-8454.2006.00274.x

Black, S. E., Lincove, J., Cullinane, J., & Veron, R. (2015). Can you leave high school behind? *Economics of Education Review, 46*, 52–63. https://doi.org/10.3386/w19842

Bonsaksen, T. (2016). Predictors of academic performance and education programme satisfaction in occupational therapy students. *British Journal of Occupational Therapy, 79*(6), 361–367. https://doi.org/10.1177/0308022615627174

Bulus, M. (2011). Goal orientations, locus of control and academic achievement in prospective teachers: An individual differences perspective. *Educational Sciences: Theory and Practice, 11*(2), 540–546.

Cascallar, E. C., Boekaerts, M., & Costigan, T. E. (2006). Assessment in the evaluation of self- regulation as a process. *Educational Psychology Review, 18*(3), 297–306. https://doi.org/10.1007/s10648-006-9023-2

Cascallar, E., Musso, M. F., Kyndt, E., & Dochy, F. (2015). Modelling for understanding and for prediction/classification the power of neural networks in research. *Frontline Learning Research, 2*(5), 67–81. https://doi.org/10.14786/flr.v2i5.135

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computers & education. *Computers & Education, 151*, 103855. https://doi.org/10.1016/j.compedu.2020.103855

Chen, X., Zou, D., & Xie, H. (2020). Fifty years of British journal of educational technology: A topic modeling based bibliometric perspective. *British Journal of Educational Technology, 51*(3), 692–708. https://doi.org/10.1111/bjet.12907

Cliffordson, C. (2008). Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment, 13*(1), 56–75.

De Clercq, M., Galand, B., Dupont, S., & Frenay, M. (2013). Achievement among first-year university students: An integrated and contextualized approach. *European Journal of Psychology of Education, 28*(3), 641–662. https://doi.org/10.1007/s10212-012-0133-6

De Clercq, M., Galand, B., & Frenay, M. (2017). Transition from high school to university: A person-centered approach to academic achievement. *European Journal of Psychology of Education, 32*(1), 39–59. https://doi.org/10.1007/s10212-016-0298-5

De Veaux, R. D., Schumi, J., Schweinsberg, J., & Ungar, L. H. (1998). Prediction intervals for neural networks via nonlinear regression. *Technometrics, 40*(4), 273–282. https://doi.org/10.2307/1270528

Fausett, L. V. (1994). *Fundamentals of neural networks*. Prentice-Hall.

Field, A. P. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)*. Sage Publications.

Garg, R. (2018). Predicting student performance in different regions of Punjab using classification techniques. *International Journal of Advanced Research in Computer Science, 9*(1), 236–241. https://doi.org/10.26483/ijarcs.v9i1.5234

Garson, G. D. (2014). *Neural network models*. Statistical Associates Publishers.

Gerken, J. T., & Volkwein, J. F. (2000, May). *Pre-college characteristics and freshman year experiences as predictors of 8-year college outcomes*. Cincinnati, OH: Paper presented at the annual meeting of the Association for Institutional Research.

Golino, H., & Gomes, C. (2014). Four machine learning methods to predict academic achievement of college students: A comparison study. *Revista E-PSI, 4*, 68–101.

Hamoud, A. K., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*(2), 26–31. https://doi.org/10.9781/ijimai.2018.02.004

Hansen, M. N., & Mastekaasa, A. (2006). Social origins and academic performance at university. *European Sociological Review, 22*(3), 277–291. https://doi.org/10.1093/esr/jci057

Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA: Pearson.

He, S., & Li, J. (2011). Confidence intervals for neural networks and applications to modeling engineering materials. In C. L. P. Hui (Ed.), *Artificial neural networks – application* (pp. 337–360). Shanghai, China: InTech.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles, and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

ICFES (Colombian Institute for the Assessment of Education). (2021a). Acerca del exámen SABER 11. [Specifications of the SABER 11 test]. http://www.icfes.gov.co/acerca-examen-saber-11.

ICFES (Colombian Institute for the Assessment of Education). (2021a). Acerca del exámen SABER PRO. [Specifications of the SABER PRO test]. http://www.icfes.gov.co/acerca-del-examen-saber-pro.

Juba, B., & Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 4039–4048. https://doi.org/10.1609/aaai.v33i01.33014039

Kewley, R., Embrechts, M., & Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks, 11*(3), 668–679. https://doi.org/10.1109/72.846738

Khalilzadeh, J., & Tasci, A. D. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management, 62*, 89–96. https://doi.org/10.1016/j.tourman.2017.03.026

Kingston, J. A., & Lyddy, F. (2013). Self-efficacy and short-term memory capacity as predictors of proportional reasoning. *Learning and Individual Differences, 26*, 185–190. https://doi.org/10.1016/j.lindif.2013.01.017

Kuncel, N. R., Crede, M., Thomas, L. L., Klieger, D. M., Seiler, S. N., & Woo, S. E. (2005). A meta-analysis of the Pharmacy College Admission Test (PCAT) and grade predictors of pharmacy student success. *American Journal of Pharmaceutical Education, 69*(3), 339–347. https://doi.org/10.5688/aj690351

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science, 315*(5815), 1080–1081. https://doi.org/10.1126/science.1136618

Kyndt, E., Musso, M., Cascallar, E. C., & Dochy, F. (2015). Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysisV. Donche, S. De Maeyer, D. Gijbels, & H. van den Bergh (Eds.). . In *Methodological challenges in research on student learning* (pp. 55–76). Garant.

Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction, and classification of student academic performance using artificial neural networks. *SN Applied Sciences, 1*, 982. https://doi.org/10.1007/s42452-019-0884-7

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 720–738. https://doi.org/10.1037/a0022639

Luft, C. D. B., Gomes, J. S., Priori, D., & Takase, E. (2013). Using online cognitive tasks to predict mathematics low school achievement. *Computers & Education, 67*, 219–228. https://doi.org/10.1016/j.compedu.2013.04.001

McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development, 20*(1), 21–33. https://doi.org/10.1080/07924360120043621

Mesarić, J., & Šebalj, D. (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review, 7*(2), 367–388. https://doi.org/10.17535/crorr.2016.0025

Mohamed, M. H., & Waguih, H. M. (2017). Early prediction of student success using a data mining classification technique. *International Journal of Science and Research, 6*(10), 126–131.

Morris, A. (2011). Student standardized testing: Current practices in OECD countries and a literature review. *OECD Education Working Papers, 65*, 1–51. https://doi.org/10.1787/5kg3rp9qbnr6-en

Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science, 11*, 36–42. https://doi.org/10.5815/ijmecs.2016.11.05

Musso, M. F. (2016). *Understanding the underpinnings of academic performance: The relationship of basic cognitive processes, self-regulation factors and learning strategies with task characteristics in the assessment and prediction of academic performance* (Publication No. LIRIAS1941186) [Doctoral dissertation, KU Leuven]. KU Leuven Digital Repository https://limo.libis.be/primo-explore/search?vid=Lirias&fromLogin=true&sortby=rank&lang=en_USC.

Musso, M. F., & Cascallar, E. C. (2009). Predictive systems using artificial neural networks: An introduction to concepts and applications in education and social sciencesM. C. Richaud, & J. E. Moreno (Eds.). . In *Research in Behavioral Sciences (Volume I)* (pp. 433–459). CIIPME/CONICET.

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2012). Predicting mathematical performance: The effects of cognitive processes and self-regulation factors. *Education Research International, 2012*, 1–13. https://doi.org/10.1155/2012/250719

Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research, 1*(1), 42–71. https://doi.org/10.14786/flr.v1i1.13

Musso, M. F., Cascallar, E. C., Bostani, N., & Crawford, M. (2020). Identifying reliable predictors of educational outcomes through machine-learning predictive modeling. *Frontiers in Education, 5*. https://doi.org/10.3389/feduc.2020.00104. Article 104.

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini review. *Frontiers in Psychology, 3*, 1–5. https://doi.org/10.3389/fpsyg.2012.00322

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research and Evaluation, 8*(2), 1–5. https://doi.org/10.7275/r222-hv23

Pérez, P. M., Costa, J. L. C., & Corbí, R. G. (2012). An explanatory model of academic achievement based on aptitudes, goal orientations, self-concept and learning strategies. *Spanish Journal of Psychology, 15*(1), 48–60. https://doi.org/10.5209/rev_sjop.2012.v15.n1.37283

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education, 43*, 187–207. https://doi.org/10.1023/A:1014419724092

Piotrowski, A. P., & Napiorkowski, J. J. (2013). A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *Journal of Hydrology, 476*, 97–111. https://doi.org/10.1016/j.jhydrol.2012.10.019

Putpuek, N., Rojanaprasert, N., Atchariyachanvanich, K., & Thamrongthanyawong, T. (2018). Comparative study of prediction models for final GPA score: A case study of Rajabhat Rajanagarindra University. *Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science, 1*, 92–97. https://doi.org/10.1109/ICIS.2018.8466475

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838

Rodríguez-Hernández, C. F., Cascallar, E., & Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review. *Educational Research Review, 29*. https://doi.org/10.1016/j.edurev.2019.100305. Article 100305.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*, 582–599. https://doi.org/10.1007/s40593-016-0110-3

Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., Shen, W., & Kiger, T. B. (2012). The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. *Psychological Science, 23*(9), 1000–1007. https://doi.org/10.1177/0956797612438732

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One, 10*(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

Shaw, E. J., Marini, J. P., Beard, J., Shmueli, D., Young, L., & Ng, H. (2016). Redesigned SAT pilot predictive validity study: A first look (Research Report No. 2016-1). https://www.collegeboard.org/.

Singh, W., & Kaur, P. (2016). Comparative analysis of classification techniques for predicting computer engineering students' academic performance. *International Journal of Advanced Research in Computer Science, 7*(6), 31–36. https://doi.org/10.26483/ijarcs.v7i6.2793

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Sivasakthi, M. (2017). Classification and prediction-based data mining algorithms to predict students' introductory programming performance. *Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI), 1*, 346–350. https://doi.org/10.1109/icici.2017.8365371

Somers, M. J., & Casal, J. C. (2009). Using artificial neural networks to model nonlinearity: The case of the job satisfaction job performance relationship. *Organizational Research Methods, 12*(3), 403–417. https://doi.org/10.1177/1094428107309326

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of MachineLearning Research, 15*(56), 1929–1958.

Steinmayr, R., Ziegler, M., & Träuble, B. (2010). Do intelligence and sustained attention interact in predicting academic achievement? *Learning and Individual Differences, 20*(1), 14–18. https://doi.org/10.1016/j.lindif.2009.10.009

Tacq, J. (1997). *Multivariate analysis techniques in social science research. From problem to analysis*. Sage Publications.

The World Bank. (2012). *Reviews of national Policies for education: Tertiary education in Colombia 2012*. OECD Publishing.

Triventi, M. (2014). Does working during higher education affect students' academic progression? *Economics of Education Review, 41*(C), 1–13. https://doi.org/10.1016/j.econedurev.2014.03.006

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*(2), 134–150. https://doi.org/10.2139/ssrn.1402645

Wang, H., Kong, M., Shan, W., & Vong, S. K. (2010). The effects of doing part-time jobs on college student academic performance and social life in a Chinese society. *Journal of Education and Work, 23*(1), 79–94. https://doi.org/10.1080/13639080903418402

Win, R., & Miller, P. W. (2005). The effects of individual and school factors on university students' academic performance. *The Australian Economic Review, 38*(1), 1–18. https://doi.org/10.1111/j.1467-8462.2005.00349.x

Yanbarisova, D. M. (2015). The effects of student employment on academic performance in Tatarstan higher education institutions. *Russian Education and Society, 57*(6), 459–482. https://doi.org/10.1080/10609393.2015.1096138

Yassein, N. A., Helali, R. G. M., & Mohomad, S. B. (2017). Information technology & software engineering predicting student academic performance in KSA using data mining techniques. *Journal of Information Technology & Software Engineering, 7*(5), 1–5. https://doi.org/10.4172/2165-7866.1000213

Yeh, I. C., & Cheng, W. L. (2010). First and second order sensitivity analysis of MLP. *Neurocomputing, 73*(10–12), 2225–2233. https://doi.org/10.1016/j.neucom.2010.01.011

Yıldız Aybek, H. S., & Okur, M. R. (2018). Predicting achievement with artificial neural networks: The case of Anadolu university open education system. *International Journal of Assessment Tools in Education, 5*(3), 474–490. https://doi.org/10.21449/ijate.435507

Zheng, J. L., Saunders, K. P., Shelley, I. I., Mack, C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development, 43*(2), 267–283.