Fall 8-7-2023

An Enhanced Artificial Neural Network Approach To Predict Student Dropout From Imbalanced Datasets

Alexandre Nascimento

Anna C. M. Queiroz

Gabriel Kenji Godoy Shimanuki

An Enhanced Artificial Neural Network Approach To Predict Student Dropout From Imbalanced Datasets

Completed Research Paper

Alexandre Moreira Nascimento
Stanford University
alexandremoreiranascimento@alum.mit.edu

Anna Carolina
Stanford University
acmq@stanford.edu

Gabriel Kenji Godoy Shimanuki
Escola Politécnica da Universidade de São Paulo
gabrielshimanuki@usp.br

Abstract

School dropout has profound and long-term impacts on global development. Machine Learning (ML) techniques have been used to create models to predict school dropouts, supporting school managers, educators, and policymakers to take proactive measures to reduce those rates and their social impact. However, most studies did not account for the imbalance of historical datasets when training those models, leading to over-optimistic performance metrics and poor practical results. In the present work, a novel ANN approach able to deal with imbalanced datasets to predict school dropout is proposed and tested with actual data, reaching the remarkable performance of correctly predicting 90.1% of the students who would dropout and outperforming the benchmark.

Keywords

school dropout; artificial intelligence; artificial neural network; dataset imbalance; tuning; optimization; sustainable development; educational data mining; EDM; ICT; ICT4D; prediction.

Introduction

Education is a crucial lever to society's development. In 2015, United Nations stated 17 sustainable development goals (SDG) for 2030 (Fritz et al., 2019), most of them highly impacted by SDG #4 "Quality education". Education impacts employability, financial gains, incarceration rates, climate change engagement, and even mental and physical health (Cutler & Lleras-Muney, 2006; T. M. Lee, Markowitz, Howe, Ko, & Leiserowitz, 2015; Orooji & Chen, 2019; Pettit & Western, 2004; Rumberger & Rotermund, 2012; Sara, Halland, Igel, & Alstrup, 2015). Thus, For that reason, UNESCO has created an international policy agenda "Education for All (EFA)" to ensure all students around the world benefit from education (S. Lee & Chung, 2019). However, most developing countries face intensive student dropouts (Mduma, Kalegele, & Machuve, 2019b), resulting in several negative social and economic consequences, and impeding their sustainable development (Del Bonifro, Gabbrielli, Lisanti, & Zingaro, 2020). According to Sara el al (2015), the financial burden of each student dropout on the rest of society in the USA is to the average of $292,000 due to lower tax income and higher incarceration, among other costs. Covid-19 pandemic worsened the perspectives in many countries, raising the school dropout rates and reducing the academic performance. According to INPER Study[1], over $38 billion was the estimated cost of school dropouts in Brazil in 2020.

[1] - https://www.wilsoncenter.org/blog-post/pandemic-worsen-brazils-troubling-school-dropout-rates

The last decade's progress in Machine Learning (ML) enabled their users to create models based on historical dropout data to predict future dropouts. Their predictions associated with proactive actions can help to reduce dropout rates and their negative consequences (Del Bonifro et al., 2020; Kostopoulos, Kotsiantis, Pierrakeas, Koutsonikos, & Gravvanis, 2018; Nagy & Molontay, 2018). Hence, ML use has been growing considerably to predict dropouts across education: K-9 education (Sara et al., 2015), high school (Chung & Lee, 2019; Gil, Delima, & Vilchez, 2020; Márquez-Vera et al., 2016), and higher education (Daza, 2022; Del Bonifro et al., 2020; Dwivedi & Pandey, 2018; Iam-On & Boongoen, 2017).

Although the importance of ML application in dropout prediction is growing, in most cases, researchers ignored the dataset imbalance (Mduma, Kalegele, & Machuve, 2019a), potentially leading to wrong predictions and decision-making. Since the fraction of dropout students is much lower than the non-dropout, the used datasets are usually imbalanced (Thammasiri, Delen, Meesad, & Kasap, 2014), making it challenging to build accurate predictive models for a dropout early detection system (S. Lee & Chung, 2019). For example, a dataset with historical data of 5% dropouts and 95% non-dropouts is highly imbalanced. When ML algorithms are trained with those unbalanced datasets, they result in very high accuracies because they learn that the best guess is to classify almost everyone as non-dropouts. Furthermore, that becomes an issue because researchers can mistakenly evaluate an 95% accuracy (or a slightly higher value) of an on artificial neural networks (ANN) as a good result while in fact it is a very poor result. By classifying 100% of the dataset as non-dropout, this algorithm, if not considering the dataset imbalance, would achieve 95% accuracy and fail to identify the 5% possible dropouts. Therefore, most systems for early detection of dropouts based on ML trained using imbalanced datasets are actually not effective. For that reason, Mduma et al. (2019a) asked the research community to develop a student dropout algorithm accounting for the data imbalance problem, which is the main goal of the present study.

Among other ML techniques (Ashfaq & Marikannan, 2019; Luan & Tsai, 2021; Mduma et al., 2019a), ANN has a solid capacity to model the complex and non-linear relationships between a target variable and independent variables (Nagy & Molontay, 2018), which makes those techniques a great candidate for dealing with complex problems such as medical diagnosis based on images (Shen, Wu, & Suk, 2017), predicting demand to reduce food waste (Nascimento, Queiroz, de Melo, & Meirelles, 2022) and dropouts (V. Kiss, Maldonado, & Segall, 2022; Mnyawami, Maziku, & Mushi, 2022; Niyogisubizo, Liao, Nziyumva, Murwanashyaka, & Nshimyumukiza, 2022). Hence, the present study proposes and tests a novel ML approach based on ANN to deal with the data imbalance problem when predicting dropouts.

The proposed approach can help to predict rare events represented by imbalanced datasets, which could be applied to other fields, such as fraud detection and customer churn prediction. The results show that the tests with the proposed approach outperformed the benchmark, reaching the remarkable performance of correctly predicting 90.1% of the students who would dropout. Those improvements in predicting dropouts can be incorporated into management information systems to support school managers and policy makers to design and implement actions to avoid dropouts, which can potentially be translated into long-term social and economic positive impacts.

The rest of this paper is organized as follows. The next section briefly describes the theoretical background on ANNs. Then, the third section briefly describes the related work on using ANNs to predict school dropouts. The fourth section presents the methodology, encompassing the proposed method, techniques, protocols, datasets, and evaluation criteria used in the present study. The fifth section reports the results obtained from the empirical evaluation and presents a discussion. Finally, the conclusions and points to future work are presented in the last section.

Artificial neural networks

ANN research started with a model of artificial neurons through mathematical abstractions of the biological neuron cell (McCulloch & Pitts, 1943). The models showed two functions governing the neuron behavior: the input function and the output/activation function (Mitchell, 1997), as illustrated in figure 1.

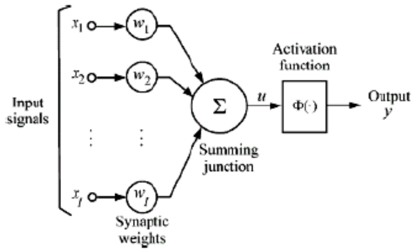


Figure 1. Diagram of an Artificial Neuron Model (Boukadida, Hassen, Gafsi, & Besbes, 2011)

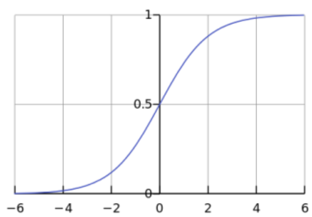
The input function is given by:

$$I_j = \sum_{i=0}^{n} w_i . x_i$$

, where the inputs X_i are weighted by the weights W_i representing the synapsis. Keeping the biological analogy, when Wi is negative, it is inhibitory because it decreases the net input; otherwise, it is excitatory (Mitchell, 1997). The output or activation function is given by:

$$y_j = f(I_j)$$

The activation function limits the permissible amplitude range of the output signal to some finite value (Karlik & Olgac, 2011). For this reason, the mathematical expressions used to implement the activation function are also known as squashing functions (Cybenko, 1989; Lippmann, 1988). Their original model has been generalized in many ways, such as using different activation functions (Jain, Mao, & Mohiuddin, 1996; Karlik & Olgac, 2011; Nair & Hinton, 2010). In these models, the sigmoid given by the logistic function is the most commonly used (Jain et al., 1996), and its equation is shown in Figure 2.



$$f(x) = \frac{1}{1 + e^{-x}}$$

Figure 2. Sigmoid Function – Adapted from Karlik and Olgac (2011)

During the 1950s and 1960s, researchers proposed many neural network models based on modifications of McCulloch e Pitts's model and Hebb`s synapses concept (Hebb, 2005). Among them was the Perceptron Learning Rule proposed by Frank Rosenblatt (Rosenblatt, 1958) to address pattern recognition problems. In Computer Science, pattern recognition is defined as a process where a pattern is given as input to a system, and the output is its classification in a specific class or category (Hebb, 2005).

While it is low effort and quick for humans to recognize patterns, they are one of the most challenging problems to be solved by machines. Before performing pattern recognition, the system needs to learn the classification during the supervised training phase, where the previously chosen and classified patterns are presented to the system. Rosenblatt (1958) originally proposed a neural network - the perceptron - encompassing three layers (Figure 3) inspired by mammals' vision system. That three layers perceptron became known as Multilayer Perceptron (MLP), which variations have become widely used until today.

One of the critical factors of the perceptron's success in the 1960s is the Perceptron Convergence Theorem (Rosenblatt, 1958) (Frank Rosenblatt, 1961). It states that if a classification problem is linearly separable, the perceptron rule of learning causes the vector of weights ω to converge to a vector, which gives a solution to the problem in a finite number of steps. However, training perceptrons were not an easy and fast task, which led to some lack of practical applications until the 1980s, when the backpropagation learning algorithm for MLP started to be used (Werbos, 1974) and reached considerable popularity with Rumelhart and colleagues' work (Rumelhart, Hinton, & Williams, 1986).

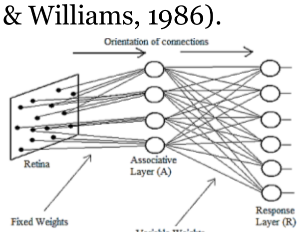


Figure 3. Rosenblatt Perceptron

The backpropagation consisted of repeatedly adjusting the weights of the connections in the network to minimize a measure of the difference between the actual output vector of the net and the desired output vector (Rumelhart et al., 1986). By doing so, the internal layers would represent essential features of the task domain, creating new valuable features and addressing some of the issues and training limitations previously found by researchers in the field.

Therefore, coupled with increased computer processing and memory power, it enabled new researches using ANNs with an increased number of layers, reducing training time from months to hours/days. ANN became popular because of its capacity to model complex and non-linear relationships among variables to predict a dependent variable, outperforming many other techniques. As computer power became cheaper and more accessible, additional layers were added to ANN inspired by the many layers of mammals' brains, and they could deal with even higher complexity. Those ANN with many layers became known as Deep Neural Networks (DNN) (an ANN type), and ML using them became known as Deep Learning (DL, a subfield of ML)(LeCun, Bengio, & Hinton, 2015).

School drop-out prediction using artificial neural networks

Using ANN for school dropout prediction is not new, since the first study found is dated to 2013 (Martinho, Nunes, & Minussi, 2013a). There are a reasonable number of studies on the topic. In fact, 24 studies were identified using *Scopus* database (Ahmed & Khan, 2019; Al-Shabandar, Hussain, Laws, Keight, & Lunn, 2017; Baranyi, Nagy, & Molontay, 2020; Daza, 2022; De Santos, Menezes, De Carvalho, & Montesco, 2019; Freitas et al., 2020; Gismondi & Huiman, 2021; Imran, Dalipi, & Kastrati, 2019; B. Kiss, Nagy, Molontay, & Csabay, 2019; V. Kiss et al., 2022; Maksimova, Pentel, & Dunajeva, 2021, 2022; Martinho et al., 2013a; Martinho, Nunes, & Minussi, 2013b; Mnyawami et al., 2022; Nagy & Molontay, 2018; Niyogisubizo et al., 2022; Nuanmeesri, Poomhiran, Chopvitayakun, & Kadmateekarun, 2022; Orooji & Chen, 2019; Park & Yoo, 2021; Sani, Nafuri, Othman, Nazri, & Nadiyah Mohamad, 2020; Segura, Mello, & Hernández, 2022; Solis, Moreira, Gonzalez, Fernandez, & Hernandez, 2018; Tsai, Chen, Shiao, Ciou, & Wu, 2020). All those studies use historical data of students' dropouts to train ANN to obtain a model that could predict future students most prone to abandon the school. However, only five studies accounted for class imbalance (Park & Yoo, 2021) (Baranyi et al., 2020) (Baranyi, Nagy, & Molontay, 2020)(Al-Shabandar et al., 2017)(Maksimova et al., 2022). That is worrisome because it indicates most studies (80%) may have presented incorrect or unrealistic conclusions about the actual performance reached and its unfolding results.

Orooji and Chen (2019) reviewed and described the techniques used to deal with unbalanced classes: *oversampling*, *undersampling*, *case weighting*, and *cost-sensitive learning*. *Oversampling* consists of techniques to fill the dataset with additional data instances of the minority classes to reduce the ratio differences among classes. Those instances can be generated by duplicating the existing instances or by interpolating two existing cases from the minority class, known by *Synthetic Minority Over-sampling Technique* (SMOTE). *Undersampling* or *downsampling* is a set of techniques to reduce the number of instances from the majority classes to smooth the ratio differences among them. In the *case weighting* approach, normalized weights are associated with the dataset instances, so the minority classes receive higher weight values than the majority classes. Therefore, the weighted fraction of minority classes becomes equivalent to that of majority classes, creating a balance. The ML algorithms consider those weights during their training, increasing the influence of the minority classes. Finally, *cost-sensitive learning* is a classification method that considers distinct costs for the misclassified instances (e.g., false positives and false negatives). Therefore, instead of minimizing the error rate, it minimizes the overall misclassification cost, compensating for dataset imbalances. Therefore, some techniques compensate for the imbalance by acting on the dataset (e.g., undersampling and oversampling). Other ones assign a weight to each instance (such as case weighting). And others act on the misclassifications by assigning a higher cost to them.

All those techniques are heavily dependent on human intervention. It requires a specialist or at least some knowledge of ML, which is not easy as shown by the proportion of scientific publications that has failed to deal with unbalanced classes. The lack of those required resources creates the risk of model results misinterpretation and may lead educators and school managers to over trust weak predictive models because of their wrong high-performance measures. Consequently, the decisions taken based on those models are not effective. As stated by Mduma et al. (2019a), there is a gap in machine learning algorithms able to deal with class imbalance, which is the goal of the present study. Algorithms able to deal with the class imbalance can reduce dependence on a more profound knowledge of ML and specialists. Moreover, they create models which can generalize the knowledge better, leading educators and school managers to more effective actions.

Methodology

This section presents the method proposed and evaluated for enhancing ANN performance when trained with imbalanced datasets.

Dataset

In the present study, a higher education (undergraduate) dropout open dataset (Martins, Tolledo, Machado, Baptista, & Realinho, 2021) from the Polytechnic Institute of Portalegre (IPP) was downloaded from *Kaggle* open online repository and used for evaluating the proposed technique. The dataset instances have ten-year data about students enrolled between the academic years 2008/2019 and 2018/2019. The data also encompasses students from eight undergraduate courses: management, education, agronomy, design, nursing, journalism, social service, and technologies. From the 4424 total instances, 1421 (32%) correspond to dropout classes, 794 (18%) to enrolled students, and 2209 (50%) to graduate students. Thus, it is an imbalanced dataset, as expected for educational institution dropout rates. Here, the complete dataset was used without the preprocessing class balancement executed by Martins et al. (2021) since the present research aims to validate the proposed approach's performance when dealing with imbalanced datasets. The dataset features are demographic, socio-economic, and academic factors. Those features are Marital status, Application mode, Application order, Course, Daytime/evening attendance, Previous qualification, Previous qualification (grade), Nationality, Mother's qualification, Father's qualification, Mother's occupation, Father's occupation, Admission grade, Displaced, Educational special needs, Debtor, Tuition fees up to date, Gender, Scholarship holder, Age at enrollment, International, Curricular units 1st sem (credited), Curricular units 1st sem (enrolled), Curricular units 1st sem (evaluations), Curricular units 1st sem (approved), Curricular units 1st sem (grade), Curricular units 1st sem (without evaluations), Curricular units 2nd sem (credited), Curricular units 2nd sem (enrolled), Curricular units 2nd sem (evaluations), Curricular units 2nd sem (approved), Curricular units 2nd sem (grade), Curricular units 2nd sem (without evaluations), Unemployment rate, Inflation rate, and GDP. An additional feature is the target class, representing the student status to be learned by the ANN: dropout, enrolled or graduated. The only data transformation performed was using the python function OrdinalEncoder() to encode the non-numeric fields and StandardScaler() to rescale the numbers from 0 to 1 for ANN training.

Two dataset versions were used. The original 3-class dataset was adopted for an experiment to evaluate the performance of the proposed approach and compare it to the literature results based on the same dataset. An additional dataset version was developed by merging the two non-dropout classes (enrolled and graduated) into a single class (non-dropout). Therefore, the new dataset version was used for the binary classification (dropout/non-dropout) experiment, the most common application found in the literature. The 3 classes in the original study also intended to support the performance evaluation, according to Martins et al. (2021), which is not the goal of the present study. Moreover, dataset binarization increased the dataset imbalance (32% dropouts/68% non-dropout), which is welcome in the present study since the goal is to evaluate the performance of the proposed technique when dealing with imbalanced datasets. Therefore, based on the binary dataset, the second experiment was intended to reproduce the most frequent situations found in the literature and provide a better understanding of the proposed model performance.

Proposed technique and experimental protocol

ANN training is an optimization problem to adjust the ANN's weights to minimize the error between its predictions and the ground truth. The proposed technique executes the ANN learning phase in two stages (training and tuning) (Figure 4). Thus, it is proposed to add a stage to the traditional ANN training phase to enhance the ANN. In the first stage, the backpropagation, a traditional ANN training stochastic gradient descendent algorithm, is used to adjust the ANN's weights until its convergence (no changes greater than a float precision equal to 2.220446049250313E-09). Thus, the first stage is based on the first-order optimization method, which was chosen here as the backpropagation. Because the ANN trained in the first stage is the traditional approach, the first stage results were considered the baseline used to compare the performance reached by the proposed approach executed in the following stage. In the second stage, a quasi-newton global optimization algorithm is used to tune the weights combined with each neuron's activation function's coefficients. Therefore, the second step's results are expected to be better since it uses more information and parameters (requires a hessian).

The ANN architecture used is a fully connected MLP (Figure 4), encompassing an input layer with an input neuron for each dataset attribute (36 neurons, except for the classification attribute), an intermediate layer with 8 neurons, and the output layer with 3 neurons corresponding to each dataset class: 3 classes (dropout, enrolled, graduated) for experiment 1; and 2 classes (dropout; non-dropout) for experiment 2. Since it is a fully connected MLP, each neuron output is connected to all the following layer's neurons. The MLP was adopted because it is widely used and requires much less computation power and energy resources than the DL. Moreover, it is compatible with the dataset size, offering a potentially good cost-benefit ratio. The learning rate was set up to 0.01. The sigmoid activation function was adopted. Beyond its frequent use in MLPs, it was selected because its non-linearity could potentially result in potentially more opportunities for tuning after the ANN's weights adjustments in the first steps.

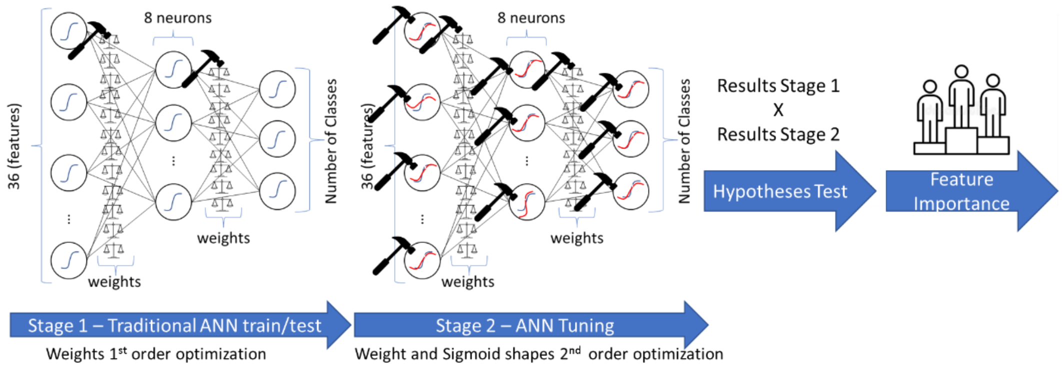


Figure 4. Experimental protocol

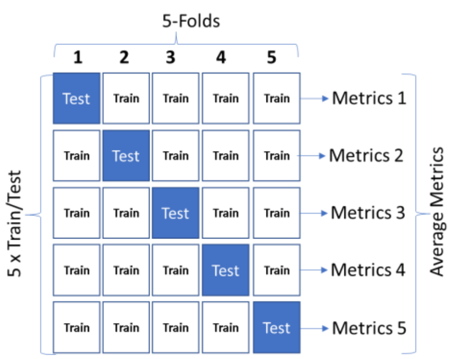


Figure 5. 5-fold cross-validation

In all experiments (one for a 3-class dataset and one for a 2-class dataset), a 5-fold cross-validation training/testing strategy was adopted to ensure a more reliable validation of the proposed technique (Figure 5). Moreover, each 5-fold cross-validation was executed 10 times. Therefore, the experiments were configured with 10 x 5-fold cross-validation. A 5-fold cross-validation means 5 training and testing are executed, and the metrics collected are averaged. The dataset is split into 5 subsets. Each ANN training is executed with 4 subsets combined and tested against the new fold. This procedure was repeated until all 5 folds were used for training and testing (but not at the same time for the same ANN) (Figure 5).

Cross-validation helps to smooth the lucky or unlucky effect when selecting the data for training and testing. The lucky effect is when a testing dataset with very easy-to-classify instances benefits the test results and leads to a false conclusion about the model performance (the same way an easy final exam could benefit unprepared students). The unfortunate effect is when a testing dataset with very hard-to-classify instances damages the test results and leads to a false conclusion about the model performance (the same way a challenging final exam could damage well-prepared students). Moreover, the approach supports a better comparison to the benchmark since Martins et al. (2021) also relied on 5-fold cross-validation training/testing.

Each experiment consisted of 50 (10 x 5-fold cross-validation) executions of the two steps, totaling 50 MLPs trained, tuned, and evaluated. At the beginning of stage 1, the MLP weights were initialized with random values (normal distribution between 0 and 1), and the 5-fold cross-validation splits were randomly drawn. Then, the trained MLP (pre-tuning) is submitted to the second stage for tuning. During the second stage, the MLP weights and each neuron's sigmoidal activation function are adjusted towards the lower sum of the squared errors until convergence. At the end of each stage, the performance evaluation metrics (subsection 0) are computed and stored. Then, right after the end of the third stage, a t-test is used to compare their values before and after ANN tuning aiming to test the hypotheses (subsection 0). Moreover, the results are compared to the benchmark results from the original study from Martins et al. (2021) based on the same dataset.

After the hypothesis tests, a feature importance analysis was performed by scaled difference in the accuracies following the same protocol used by Chung and Lee (2019). Since two experiments were performed (based on distinct dataset versions), the feature importance analysis was performed for each experiment setup. The analysis was performed by removing each one of the features at a time and executing the proposed approach to compute the results, using the same configuration and dataset splits for 5-fold cross-validation. After the steps described, the removed feature is reinserted into the dataset, and another feature is removed. Then, those steps are re-executed until all the features have been removed once, and the induced ANN without that feature has its performance metrics evaluated. Then, each result for each missing feature is compared to the benchmark (ANN trained and tuned using all the features). The feature that causes the most significant drop in each performance metric is considered the most important feature. The ratio of each performance drop caused by each of the remaining features over the most important feature is computed. Those ratios are the relative importance of each feature compared to the most important one. Those values are used to sort the feature importance per performance metric.

The experimental protocol was implemented on a python 3.6.13 code in a Jupyter Notebook. The second stage for tuning the sigmoid shapes was implemented using the BFGS optimizer from the Python library scipy 1.5.4.

Research question, performance metrics, and hypotheses

The present study's main research question is: "Is it possible to improve the network obtained in conventional training based on an imbalanced dropout dataset using a combined optimization of weights and activation functions?". Eleven metrics were used to evaluate the proposed technique's performance when dealing with an imbalanced dropout dataset. **True positives (TP)** are the total number of instances the ANN classified correctly as dropouts when executing the test dataset classification. Thus, the higher the TP, the better the ANN. **True negatives (TN)** are the total number of instances from the test dataset the ANN classified correctly as non-dropouts. Then, the higher the TN, the better the ANN. **False positives (FP)** are the total number of instances from the test dataset misclassified as dropouts but non-dropouts. Thus, minimizing FP is desired since they lead educators and school management to drive time and resources to students who would not dropout. **False negatives (FN)** are the total number of instances misclassified as non-dropouts, which were non-dropouts. Thus, minimizing FN is also desired since they divert the educators' and managers' attention and resources from students who will dropout. **Accuracy (Acc)** is the total number of instances classified correctly, which is given by (TP+TN)/(TP+TN+FP+FN). Thus, the higher the Acc, the better the predictive model. **Recall (Rcl)** is the total number of actual dropouts classified correctly (TP) from the total number of instances the model classified as dropouts (TP + FN). R is given by TP/(TP+FN). Higher R, the better the classification model too. **Precision (P)** measures the accuracy of instances classified as positive: the ratio of instances classified as positive that were positive classes. P is given by TP/(TP+FP), and the higher its value, the better the model. **F1-score (F)** is the geometric mean of P and R, given by (2xPxR)/(P+R). That is, it helps to evaluate both metrics in a combined way since there is a natural trade-off between P and R when the machine learning model (ANN in this study) choose the threshold to determine if a class is positive or negative. As the ANN is being trained, it adjusts the threshold, and it may cause the R to increase and the P to decrease or vice versa. Since F helps evaluate their combination, the higher its value, the better the model. The area under the Precision-Recall curve (AUC-PR) is the area under the curve plotting the P x R trade-off values. Thus, the higher AUC-PR, the better the ANN is. AUC-PR is an excellent metric for evaluating imbalanced datasets, better than **AUC-ROC** (Saito & Rehmsmeier, 2015; Sofaer, Hoeting, & Jarnevich, 2019). During the ANN training process, the machine learning algorithm seeks to minimize the **loss (L)** function, which was chosen as the sum of the squared error (between ground truth and prediction). The lower the L value achieved with the classification of the testing dataset, the better the ANN.

The following hypothesis were formulated based on the expected effect of the proposed ANN tuning approach on those metrics, and they were tested using the experimental protocol described in the next section to answer the research question: H1 – Tunning trained ANN by optimizing its sigmoid activation function increases its TP; H2 – Tunning trained ANN by optimizing its sigmoid activation function increases its TN; H3 – Tunning trained ANN by optimizing its sigmoid activation function reduces its FP; H4 - Tunning trained ANN by optimizing its sigmoid activation function reduces its FN; H5 - Tunning trained ANN by optimizing its sigmoid activation function increases its Acc; H6 - Tunning trained ANN by optimizing its sigmoid activation function increases its R; H7 - Tunning trained ANN by optimizing its sigmoid activation function increases its P; H8 - Tunning trained ANN by optimizing its sigmoid activation function increases its F; H9 - Tunning trained ANN by optimizing its sigmoid activation function reduces its L; H10 - Tunning trained ANN by optimizing its sigmoid activation function increases its AUC-PR; H11 - Tunning trained ANN by optimizing its sigmoid activation function increases its AUC-ROC.

Results and discussion

Table 3 shows the experimental results based on the original dataset (3-classes), which was executed to test the hypotheses and answer the research question. Each performance metric evaluated and their values after the regular ANN train (pre-tuning) and after the tunning (post-tuning) are also listed in the table. Finally, Table 3 presents the improvements from the proposed technique, the statistical significance from the t-test, and the acceptance (or not) of each hypothesis. Results showed that the proposed approach significantly improved prediction accuracy in all the metrics except for true positives. The 16.7% reduction in false positives and its standard deviation indicated that this approach could reduce the resources allocated to help students not risk dropping out. Moreover, the reduction of 5.3% of false negatives shows a potential reduction in the number of students that otherwise would not receive any support to reduce their dropout intention because of model misclassifications.

The proposed approach increased the average Acc of the models by 3.9% and reduced its standard deviation with statistical significance (p-value < 0.01%). Thus, H5 was accepted. Acc improvement is an important result for better decision-making support. Moreover, the Acc was improved with the same dataset. The results achieved by Martins et al. (2021) ranged from 60% to 73%, with the best results reached with boosting models (which usually outperform traditional models) and dataset balancing approaches. Thus, the proposed approach demonstrated superior results, with accuracies ranging from 73.2% to 79% after the ANN tuning. Since the results achieved by the MLP before the ANN tuning also exceed those obtained by Martins et al. (2021) (who did not use ANN), it is reasonable to state the ANN can outperform the other approaches and that result can be enhanced with the proposed approach.

The proposed approach improved the average R and P by 3.9% and 22.4%, respectively. Their improvements were statistically significant at 0.01% level. Thus, H6 and H7 were accepted. P improvement was considerable, which is highly desired in dropout prediction or other rare events. F was also improved considerably by 13.4% at a 0.01% significance level, supporting the acceptance of H8. The proposed approach reduced the average L by 27,5% at a 0.01% significance level, supporting H9 acceptance. Finally, AUC-PR is an important metric to evaluate the model's performance, especially for models trained with imbalanced datasets. The proposed approach enhanced the AUC-PR by 7,2% at a 0.01% significance level, supporting H10 acceptance.

		Trained ANN (10 x 5-fold cross validation)										
		Traditional Train (pre-tuning)				Proposed approach (post-tuning)						
Hn	Performance Metric	Average	Std. Dev	Min	Max	Average	Std. Dev	Min	Max	Improv.	Sig.	Result
H1	True Positive (TP)	285.9	10.0	255.6	304.6	288.8	9.6	259.9	307.0	1.0%		Rejected
H2	True Negative (TN)	443.4	5.3	431.7	455.5	460.2	5.6	449.8	472.0	3.8%	***	Accepted
H3	False Positive (FP)	100.5	4.5	91.2	111.0	83.7	4.1	75.9	94.0	16.7%	***	Accepted
H4	False Negative (FN)	55.0	4.2	45.1	67.9	52.1	3.7	44.6	63.5	5.3%	***	Accepted
H5	Accuracy (Acc)	74.0%	1.3%	70.2%	77.1%	76.9%	1.2%	73.2%	79.0%	3.9%	***	Accepted
H6	Recall (R)	74.5%	0.8%	72.7%	76.2%	77.4%	0.7%	75.7%	79.2%	3.9%	***	Accepted
H7	Precision (P)	61.1%	1.1%	57.9%	63.6%	74.8%	1.5%	70.1%	77.5%	22.4%	***	Accepted
H8	F-Measure (F)	67.1%	0.9%	64.7%	69.3%	76.1%	1.1%	73.0%	78.1%	13.4%	***	Accepted
H9	Loss (L)	25.8%	0.8%	24.6%	29.6%	18.7%	0.7%	17.1%	20.3%	27.5%	***	Accepted
H10	Area under curve PR (AUC-PR)	75.9%	1.5%	72.5%	80.0%	81.4%	1.4%	78.3%	84.4%	7.2%	***	Accepted
H11	Area under curve ROC (AUC-ROC)	85.5%	1.1%	83.1%	88.3%	89.8%	0.8%	87.9%	91.1%	5.0%	***	Accepted

** 5%, ** 0.1%, ** 0.01%*

Table 1. Experiment 1 (3 classes) results and hypothesis tests

		Trained ANN (10 x 5-fold cross validation)										
		Traditional Train (pre-tuning)				Proposed approach (post-tuning)						
Hn	True Positive (TP)	Average	Std. Dev	Min	Max	Average	Std. Dev	Min	Max	Improv.	Sig.	Result
H1	True Positive (TP)	576.0	13.3	543.0	614.0	569.5	14.9	536.0	602.0	-1%	*	Rejected
H2	True Negative (TN)	188.2	13.4	153.0	212.0	209.5	14.5	177.0	237.0	11%	***	Accepted
H3	False Positive (FP)	96.0	12.6	72.0	154.0	74.7	8.3	55.0	97.0	22%	***	Accepted
H4	False Negative (FN)	24.6	6.1	11.0	46.0	31.1	4.9	19.0	42.0	-27%	***	Rejected
H5	Accuracy (Acc)	86.4%	1.3%	81.4%	89.0%	88.0%	1.0%	85.9%	90.2%	2%	***	Accepted
H6	Recall (R)	85.8%	1.3%	81.7%	88.8%	87.6%	1.0%	85.7%	90.0%	2%	***	Accepted
H7	Precision (P)	86.7%	1.2%	83.6%	89.4%	88.0%	1.0%	85.9%	90.1%	1%	***	Accepted
H8	F-Measure (F)	86.2%	1.2%	82.6%	89.1%	87.8%	1.0%	85.9%	90.0%	2%	***	Accepted
H9	Loss (L)	17.0%	0.7%	15.8%	20.4%	11.2%	0.8%	9.8%	13.5%	34%	***	Accepted
H10	Area under curve PR (AUC-PR)	91.4%	0.9%	89.6%	93.5%	92.6%	0.8%	90.9%	94.2%	1%	***	Accepted
H11	Area under curve ROC (AUC-ROC)	90.6%	1.0%	88.4%	93.1%	91.9%	0.9%	89.9%	93.6%	1%	***	Accepted

** 5%, ** 0.1%, ** 0.01%*

Table 2. Experiment 2 (2 classes) results and hypothesis tests

It is important to mention that despite the significance, the average improvement in TP was observed, and all other metrics were improved with a very high confidence interval. Thus, it is possible to improve the network obtained in conventional training based on imbalanced dropout dataset using a combined optimization of weights and activation functions. The improvements were observed under the same condition in a controlled setup described in the present study and when compared to the original study (Martins et al., 2021) based on the same dataset.

Experiment 2 was executed to evaluate the proposed approach in a more frequent setup found in the literature. Table 4 shows the results for experiment 2 (analogously to Table 3) based on the binary dataset version. Contrary to what was expected, TP was reduced by 1%. Furthermore, FN increased by 27%, rejecting H1 and H4, respectively. Since the results achieved by the MLP before the ANN tuning also exceed those obtained by Martins et al. (2021) (who did not use ANN), it is reasonable to state the ANN can outperform the other approaches and that result can be enhanced with the proposed approach.

The proposed approach improved the average R and P by 2% and 1%, respectively. Their improvements were statistically significant at 0.01% level. R reached a maximum value of 90% and P reached a value of 90.1%, which are great performance levels and impressive for imbalanced datasets. Thus, H6 and H7 were accepted. F also reached 90% and was improved by 2% at a 0.01% significance level, supporting the acceptance of H8. The proposed approach reduced the average L by 34% at a 0.01% significance level, supporting H9 acceptance. The loss reduction was remarkable. AUC-PR was improved by 1%, reaching 94.2% at a 0.01% significance level, supporting H10 acceptance. That was another remarkable achievement.

Finally, AUC-ROC was also improved by 1% at a 0.01% significance level, supporting H11. Therefore, except for H1 and H4, all other hypotheses were accepted, supporting an affirmative answer to the research question under the same arguments presented previously for experiment 1.

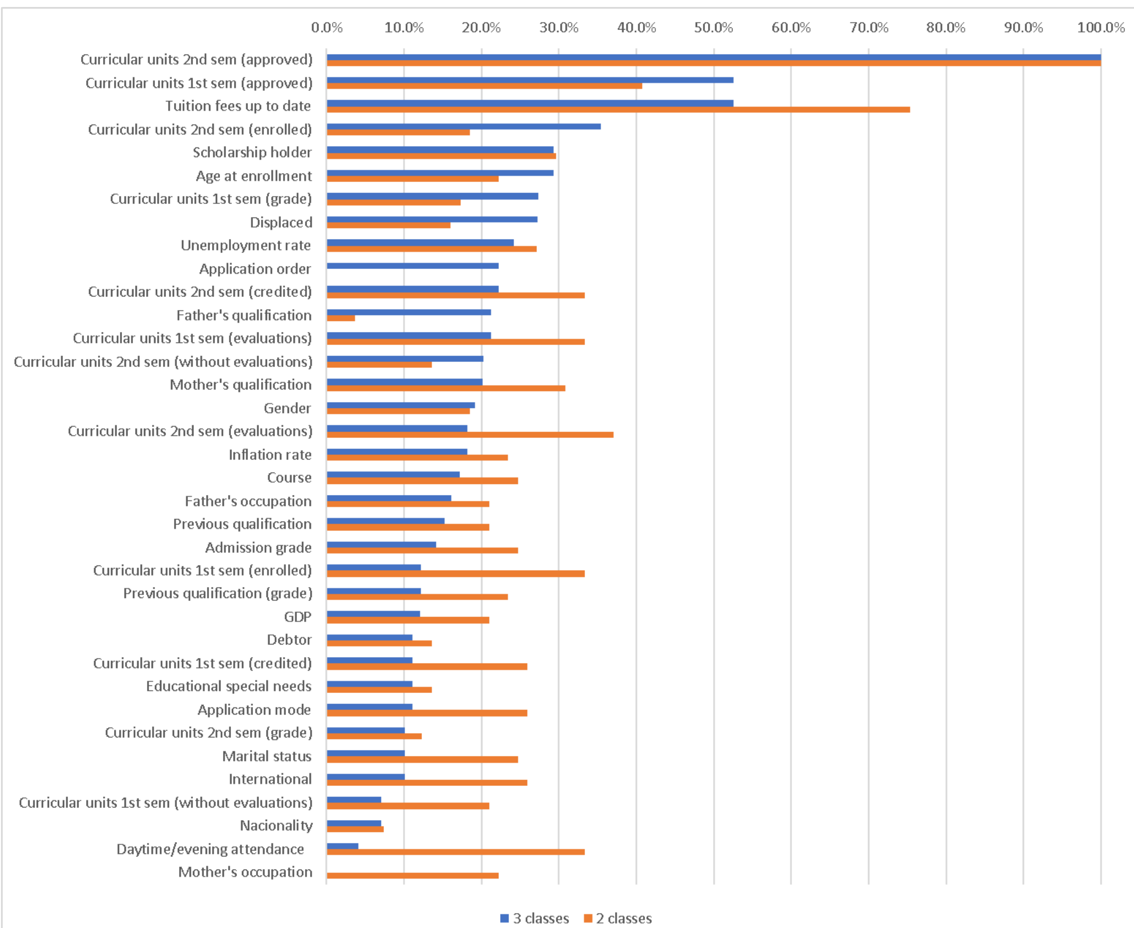


Figure 6. Feature importance ranking for each dataset

Figure 6 shows the importance of each feature relative to the most important one for ANN created with 3-classes dataset (experiment 1) and 2-classes dataset (experiment 2). The analysis can help to identify the factors that highly affect student dropout to help to plan actions (Gil et al., 2020). For both setups, the most important feature was the *Curricular units 2nd the semester (approved)*, which is aligned with Gil et al. (2020), who stated that academic performance is the top indicator for students' dropout cases. The other two features at the same position at the importance ranking for both experiments are Inflation rate and Nationality, where the first one is more important than the second for both setups. The most important feature in predicting the student dropout is related to the student's academic performance and it persisted for both experiments, where the first one predicted dropout and also the graduation (which is related academic performance), and the second one focused only on dropout prediction. The other common features at the same importance position is an economic factor (*Inflation rate*) and a demographic factor (*Nationality)*. Thus, despite of the importance of socio-economic and demographic variables, which are harder to be managed, the academic performance influence in school dropout demonstrates there are possible preventive actions to be taken by educators and school managers to attenuate the number of students abandoning the schools.

Conclusion

Student dropout from educational institutions have profound personal, social and economic consequences. Researchers have been using ML techniques, such as ANN to predict dropouts, so proactive measures can be taken to reduce them. However, most research found (80%) did not take into account the side effects on those predictions when imbalanced datasets are used, as pointed out by Mduma et al. (2019a) and confirmed in the present study. The main goal of the present study was to propose and test an ANN approach to deal automatically with imbalanced datasets.

Two dropout dataset versions were used to test 11 hypotheses based on the performance metrics used to evaluate the proposed approach without using any dataset balance technique. Achieved results demonstrated that the approach enhanced ANN, which was able to correctly predict over 90% of the students who would dropout, outperforming the original study's results (Martins et al., 2021) based on the same data. If implemented correctly, the proposed approach can positively impact student retention, helping school managers and educators proactively and give early additional support to students at risk of dropping out. That impact could unfold long-term societal positive impacts and help the achievement of the UN's Sustainable Development Goals.

Additionally, a feature importance analysis was performed. Those results pointed out the importance of the previous academic performance in the school dropout. Thus, despite the importance of socio-economic and demographic variables, which are harder to control, actions to improve students' academic achievement can potentially reduce the dropout rate.

The present research has limitations which will be addressed by future research. Only one ANN architecture and a dataset (two versions) were evaluated. Moreover, the approach was tested only with higher education institutions from a country. Therefore, future research must evaluate the proposed approach using many distinct ANN architectures, configurations, and datasets from distinct school levels and countries. Furthermore, the proposed technique will be evaluated for other societal and business applications with similar data imbalance issues related to predicting low-frequency events, such as frauds, customer churn, and credit approval based on historical data.

References

Ahmed, S. A., & Khan, S. I. (2019). A machine learning approach to Predict the Engineering Students at risk of dropout and factors behind: Bangladesh Perspective. In *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*. https://doi.org/10.1109/ICCCNT45670.2019.8944511

Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., & Lunn, J. (2017). Towards the differentiation of initial and final retention in massive open online courses. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10361 LNCS*, 26–36. https://doi.org/10.1007/978-3-319-63309-1_3

Ashfaq, U., & Marikannan, B. P. (2019). A survey: A predictive analytics approach using imbalanced data. *International Journal of Pharmaceutical Research*, *11*, 449–454. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079245835&partnerID=40&md5=fd9e3f9d0eff6ae7ea0c49861ebca0f9

Baranyi, M., Nagy, M., & Molontay, R. (2020). Interpretable Deep Learning for University Dropout Prediction. In *SIGITE 2020 - Proceedings of the 21st Annual Conference on Information Technology Education* (pp. 13–19). https://doi.org/10.1145/3368308.3415382

Boukadida, H., Hassen, N., Gafsi, Z., & Besbes, K. (2011). A HIGHLY TIME-EFFICIENT DIGITAL MULTIPLIER BASED ON THE A2 BINARY REPRESENTATION. *International Journal of Engineering Science and Technology*, *3*.

Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, *96*, 346–353.

Cutler, D. M., & Lleras-Muney, A. (2006). Education and health: evaluating theories and evidence. National bureau of economic research Cambridge, Mass., USA.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314.

Daza, A. (2022). A STACKING BASED HYBRID TECHNIQUE TO PREDICT STUDENT DROPOUT AT UNIVERSITIES. *Journal of Theoretical and Applied Information Technology*, *100*(13), 4790–4801. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134385654&partnerID=40&md5=f313839883377ee32831a02f6755ad2a

De Santos, K. J. O., Menezes, A. G., De Carvalho, A. B., & Montesco, C. A. E. (2019). Supervised learning in the context of educational data mining to avoid university students dropout. In *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019* (pp. 207–208). https://doi.org/10.1109/ICALT.2019.00068

Del Bonifro, F., Gabbrielli, M., Lisanti, G., & Zingaro, S. P. (2020). Student dropout prediction. In *International Conference on Artificial Intelligence in Education* (pp. 129–140).

Dwivedi, S. K., & Pandey, P. (2018). Data Mining In Higher Education: Students Dropout Assessment "A Case Study on Aps University." *Journal of Advances and Scholarly Researches in Allied Education [JASRAE]*, *15*(9), 157-164 (8). Retrieved from http://ipublisher.in/I/a/57970

Freitas, F. A. D. S., Vasconcelos, F. F. X., Peixoto, S. A., Hassan, M. M., Ali Akber Dewan, M., de Albuquerque, V. H. C., & Rebouças Filho, P. P. (2020). IoT system for school dropout prediction using machine learning techniques based on socioeconomic data. *Electronics (Switzerland)*, *9*(10), 1–14. https://doi.org/10.3390/electronics9101613

Fritz, S., See, L., Carlson, T., Haklay, M. M., Oliver, J. L., Fraisl, D., … others. (2019). Citizen science and the United Nations sustainable development goals. *Nature Sustainability*, *2*(10), 922–930.

Gil, J. S., Delima, A. J. P., & Vilchez, R. N. (2020). Predicting students' dropout indicators in public school using data mining approaches. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(1), 774–778.

Gismondi, H. E. C., & Huiman, L. V. U. (2021). Multilayer Neural Networks for Predicting Academic Dropout at the National University of Santa - Peru. In *Proceedings - 7th International Symposium on Accreditation of Engineering and Computing Education, ICACIT 2021*. https://doi.org/10.1109/ICACIT53544.2021.9612507

Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Iam-On, N., & Boongoen, T. (2017). Generating descriptive model for student dropout: a review of clustering approach. *Human-Centric Computing and Information Sciences*, *7*(1), 1–24.

Imran, A. S., Dalipi, F., & Kastrati, Z. (2019). Predicting student dropout in a MOOC: An evaluation of a deep neural network model. In *ACM International Conference Proceeding Series* (pp. 190–195). https://doi.org/10.1145/3330482.3330514

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, *29*(3), 31–44.

Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, *1*(4), 111–122.

Kiss, B., Nagy, M., Molontay, R., & Csabay, B. (2019). Predicting dropout using high school and first-semester academic achievement measures. In *ICETA 2019 - 17th IEEE International Conference on Emerging eLearning Technologies and Applications, Proceedings* (pp. 383–389). https://doi.org/10.1109/ICETA48886.2019.9040158

Kiss, V., Maldonado, E., & Segall, M. (2022). The Use of Semester Course Data for Machine Learning Prediction of College Dropout Rates. *Journal of Higher Education Theory and Practice*, *22*(4), 64–74. https://doi.org/10.33423/jhetp.v22i4.5133

Kostopoulos, G., Kotsiantis, S., Pierrakeas, C., Koutsonikos, G., & Gravvanis, G. A. (2018). Forecasting students' success in an open university. *International Journal of Learning Technology*, *13*(1), 26–43. https://doi.org/10.1504/IJLT.2018.091630

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, S., & Chung, J. Y. (2019). The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences*, *9*(15), 3093.

Lee, T. M., Markowitz, E. M., Howe, P. D., Ko, C.-Y., & Leiserowitz, A. A. (2015). Predictors of public climate change awareness and risk perception around the world. *Nature Climate Change*, *5*(11), 1014–1020.

Lippmann, R. P. (1988). An Introduction to Computing with Neural Nets. *SIGARCH Comput. Archit. News*, *16*(1), 7–25. https://doi.org/10.1145/44571.44572

Luan, H., & Tsai, C.-C. (2021). A Review of Using Machine Learning Approaches for Precision Education. *Educational Technology and Society*, *24*(1), 250–266. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102828171&partnerID=40&md5=019cf02a1178a0511a111961c3c68fd0

Maksimova, N., Pentel, A., & Dunajeva, O. (2021). Predicting First-Year Computer Science Students Drop-Out with Machine Learning Methods: A Case Study. *Advances in Intelligent Systems and Computing*, *1329*, 719–726. https://doi.org/10.1007/978-3-030-68201-9_70

Maksimova, N., Pentel, A., & Dunajeva, O. (2022). Computer Science Students Early Drop-Out Prediction Using Machine Learning: A Case Study. *Lecture Notes in Networks and Systems*, *456*, 523–549. https://doi.org/10.1007/978-3-031-04286-7_25

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: a case study with high school students. *Expert Systems*, *33*(1), 107–124.

Martinho, V. R. C., Nunes, C., & Minussi, C. R. (2013a). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. In *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI* (pp. 159–166). https://doi.org/10.1109/ICTAI.2013.33

Martinho, V. R. C., Nunes, C., & Minussi, C. R. (2013b). Prediction of school dropout risk group using neural network. In *2013 Federated Conference on Computer Science and Information Systems, FedCSIS 2013* (pp. 111–114). Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84892496898&partnerID=40&md5=fe290ed06c0eae6775e968f27d93483c

Martins, M. V, Tolledo, D., Machado, J., Baptista, L. M. T., & Realinho, V. (2021). Early Prediction of student's Performance in Higher Education: A Case Study. In *World Conference on Information Systems and Technologies* (pp. 166–175).

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

Mduma, N., Kalegele, K., & Machuve, D. (2019a). A survey of machine learning approaches and techniques for student dropout prediction.

Mduma, N., Kalegele, K., & Machuve, D. (2019b). Machine learning approach for reducing students dropout rates.

Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

Mnyawami, Y. N., Maziku, H. H., & Mushi, J. C. (2022). Enhanced Model for Predicting Student Dropouts in Developing Countries Using Automated Machine Learning Approach: A Case of Tanzanian's Secondary Schools. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2022.2071406

Nagy, M., & Molontay, R. (2018). Predicting Dropout in Higher Education Based on Secondary School Performance. In *INES 2018 - IEEE 22nd International Conference on Intelligent Engineering Systems, Proceedings* (pp. 389–394). https://doi.org/10.1109/INES.2018.8523888

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807–814).

Nascimento, A. M., Queiroz, A., de Melo, V. V, & Meirelles, F. S. (2022). Applying Artificial Intelligence to Reduce Food Waste in Small Grocery Stores.

Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, *3*. https://doi.org/10.1016/j.caeai.2022.100066

Nuanmeesri, S., Poomhiran, L., Chopvitayakun, S., & Kadmateekarun, P. (2022). Improving Dropout Forecasting during the COVID-19 Pandemic through Feature Selection and Multilayer Perceptron Neural Network. *International Journal of Information and Education Technology*, *12*(9), 851–857. https://doi.org/10.18178/ijiet.2022.12.9.1693

Orooji, M., & Chen, J. (2019). Predicting Louisiana Public High School Dropout through Imbalanced Learning Techniques. In *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 456–461).

Park, H. S., & Yoo, S. J. (2021). Early Dropout Prediction in Online Learning of University using Machine Learning. *International Journal on Informatics Visualization*, *5*(4), 347–353. https://doi.org/10.30630/JOIV.5.4.732

Pettit, B., & Western, B. (2004). Mass imprisonment and the life course: Race and class inequality in US incarceration. *American Sociological Review*, *69*(2), 151–169.

Rosenblatt, F. (1958). The {P}erceptron: {A} probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*, 386–408.

Rosenblatt, Frank. (1961). *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*.

Rumberger, R. W., & Rotermund, S. (2012). The relationship between engagement and high school dropout. In *Handbook of research on student engagement* (pp. 491–513). Springer.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533. Retrieved from http://dx.doi.org/10.1038/323533a0

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), e0118432.

Sani, N. S., Nafuri, A. F. M., Othman, Z. A., Nazri, M. Z. A., & Nadiyah Mohamad, K. (2020). Drop-Out Prediction in Higher Education Among B40 Students. *International Journal of Advanced Computer Science and Applications*, *11*(11), 550–559. https://doi.org/10.14569/IJACSA.2020.0111169

Sara, N.-B., Halland, R., Igel, C., & Alstrup, S. (2015). High-school dropout prediction using machine learning: A Danish large-scale study. In *ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence* (pp. 319–324).

Segura, M., Mello, J., & Hernández, A. (2022). Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics*, *10*(18). https://doi.org/10.3390/math10183359

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, *19*, 221.

Sofaer, H. R., Hoeting, J. A., & Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, *10*(4), 565–577. https://doi.org/https://doi.org/10.1111/2041-210X.13140

Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018). Perspectives to Predict Dropout in University Students with Machine Learning. In *2018 IEEE International Work Conference on Bioinspired Intelligence, IWOBI 2018 - Proceedings*. https://doi.org/10.1109/IWOBI.2018.8464191

Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, *41*(2), 321–330. https://doi.org/https://doi.org/10.1016/j.eswa.2013.07.046

Tsai, S.-C., Chen, C.-H., Shiao, Y.-T., Ciou, J.-S., & Wu, T.-N. (2020). Precision education with statistical learning and deep learning: a case study in Taiwan. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00186-2

Werbos, P. (1974). Beyond Regression:" New Tools for Prediction and Analysis in the Behavioral Sciences. *Ph. D. Dissertation, Harvard University*.