

Multidisciplinary Journal of Educational Research
Online First – First Published on 5th June, 2024
© The Author(s) 2024
<http://dx.doi.org/10.17583/remie.13342>

Effectiveness of Artificial Intelligence Models for Predicting School Dropout: A Meta-Analysis

Jacobo Roda-Segarra¹, Cristina de-la-Peña², & Santiago Mengual-Andrés¹

1) *Universitat de València, Spain*

2) *Universidad Internacional de La Rioja, Spain*

Abstract

School dropout is a major concern in the educational systems of all countries. In recent years, artificial intelligence is playing an important role in predicting school dropout in the different educational stages of formal education. In this context, it is crucial to know that these predictions are accurate and understandable. This meta-analytic study aims to investigate the effectiveness of dropout prediction models conducted until May 2022. The databases used are Web of Science, Scopus, PubMed, ERIC, PsyInfo, Dialnet and Scielo. 15 studies with a sample size of 199,015 participants are analyzed. The meta-analysis uses a random-effects proportions model with 95% confidence interval. Statistical evidence indicates that artificial intelligence models performed well (91%; 95% CI = 89-93%) in predicting dropout; specifically, the Decision Tree model significantly (95.3%; 95% CI = 93-98%) predicts dropout better than other models such as Random Forest, Artificial Neural Network, Support Vector Machines, Logistic Regression and Stacking Ensemble. Consequently, more models should be applied in the dropout field with larger numbers of participants to confirm these findings and improve the quality of education.

Keywords

Artificial intelligence, dropout, formal education, meta-analysis

To cite this article: Roda-Segarra, J., de-la-Peña, C., & Mengual-Andrés, S. (2024). Effectiveness of artificial intelligence models for predicting school dropout: a meta-analysis. *Multidisciplinary Journal of Educational Research*, Online First – First Published on 5th June, 2024, pp. 1-24. <http://dx.doi.org/10.17583/remie.13342>

Corresponding author(s): Jacobo Roda-Segarra

Contact address: jacobo.roda@uv.es

Multidisciplinary Journal of Educational Research
Online First – Primera publicación el 5 de Junio de 2024
© Autor(s) 2024
<http://dx.doi.org/10.17583/remie.13342>

Eficacia de los Modelos de Inteligencia Artificial para Predecir el Abandono Escolar: Un Meta-Análisis

Jacobo Roda-Segarra¹, Cristina de-la-Peña², y Santiago Mengual-Andrés¹

1) *Universitat de València*, España

2) *Universidad Internacional de La Rioja*, España

Resumen

El abandono escolar es una de las principales preocupaciones de los sistemas educativos de todos los países. En los últimos años, la inteligencia artificial está desempeñando un papel importante en la predicción del abandono escolar en las diferentes etapas educativas de la educación formal. En este contexto, es crucial saber si estas predicciones son precisas y comprensibles. Este estudio meta-analítico tiene como objetivo investigar la eficacia de los modelos de predicción del abandono escolar realizados hasta mayo de 2022. Las bases de datos utilizadas son Web of Science, Scopus, PubMed, ERIC, PsyInfo, Dialnet y Scielo. Se analizan 15 estudios con un tamaño muestral de 199.015 participantes. El metaanálisis utiliza un modelo de proporciones de efectos aleatorios con un intervalo de confianza del 95%. Las pruebas estadísticas indican que los modelos de inteligencia artificial obtuvieron buenos resultados (91%; IC del 95% = 89-93%) en la predicción del abandono; concretamente, el modelo de árbol de decisión predice significativamente (95,3%; IC del 95% = 93-98%) mejor el abandono que otros modelos como Random Forest, Artificial Neural Network, Support Vector Machines, Logistic Regression y Stacking Ensemble. En consecuencia, deberían aplicarse más modelos en el campo del abandono con un mayor número de participantes para confirmar estos resultados y mejorar la calidad de la educación.

Palabras clave

Inteligencia artificial, abandono, educación formal, meta-análisis

Cómo citar este artículo: Roda-Segarra, J., de-la-Peña, C., y Mengual-Andrés, S. (2024). Effectiveness of artificial intelligence models for predicting school dropout: a meta-analysis. *Multidisciplinary Journal of Educational Research*, Online First – Primera publicación el 5 de Junio de 2024, pp. 1-24. <http://dx.doi.org/10.17583/remie.13342>

Correspondencia Autores(s): Jacobo Roda-Segarra

Dirección de contacto: jacobo.roda@uv.es

Artificial intelligence (AI) used for prediction is an ever-expanding field of current relevance for creating both early detection and personalized recommendation systems that optimize the teaching-learning process and educational quality (Du et al., 2020; Xiao et al., 2018). According to the 2019 Horizon Report (Educause, 2019) the application of AI to educational processes will grow by more than 43%.

AI is a diverse set of methods and technologies (data mining, neural networks, algorithms, etc.) that allow computers to simulate tasks generally associated to a human mind (Baker & Smith, 2019). With this technology, computers should offer answers no worse than those that a human being could provide (Dobrev, 2012). Together with Turing's thesis, which states that what is relevant is not the fact that a machine may think or not, but that it really seems to think, AI is reduced to a behaviorist process based on imitations (Mira et al., 2003). These authors argue that, in order for a technology to be defined as AI, learning must take place.

This learning occurs through a series of algorithms, which are abstract machines that follow a sequence of steps to obtain results (Moschovakis, 2001) and, additionally in the case of supervised AIs, through a training process. In this process, the AI is provided with a series of input data, as well as the response it is expected to provide and, through the steps indicated in its algorithm, its internal structure is modified to provide the expected results. In this way, it can provide appropriate outputs for new input data for which its output is unknown. This is defined as generalization (Aldabas-Rubira, 2002).

One of the methods of AI is machine learning (Popenici & Kerr, 2017) which facilitates the creation of, among other things, predictions from data, such as school dropout rates.

School dropout refers to excused or unexcused absences from classes at any stage of education (Kearney, 2008) and completion in the educational system without a degree (Mahoney, 2018). This paper focuses on non-degree completion as an unresolved problem in the educational system (Kim & Kim, 2018).

In the context of students dropping out of school in the different educational stages that conform formal education, the relevance of AI is important for educational systems in all countries (principals, managers and politicians). School dropout is one of the most concerning issues for educational policies, especially after the implications of the measures taken to contain the COVID-19 pandemic with the closure of schools and the change from face-to-face to distance learning (Day et al., 2021). In this sense, Reuge et al. (2021), within the framework of a research proposed by UNICEF, explain how students who belong to groups at risk of social exclusion drop out of school permanently, aggravated by the impact of the measures taken to stop the spread of the COVID-19 pandemic. However, belonging to a group at risk of social exclusion is not the only risk factor for dropping out of school. Other circumstances, such as individual factors, or those of the educational systems themselves, which are not able to respond to this problem, must also be taken into account (UNICEF, 2017). Studies on school dropout (Heublein, 2014; Wilcoxson et al., 2011) identify several factors involved such as academic performance, cognitive potential, economic limitation, lack of motivation, study habits, personal circumstances, failure in the choice of study or wrong expectation. In fact, UNICEF itself proposes a warning system based on a series of indicators (Early Warning System) to detect the risk of dropping out of school in contexts where dropout exceeds 10% of total amount of students (UNICEF, 2018).

Educational institutions can take action to reduce the dropout rate, which leads to the professional, economic and social impoverishment of countries. These actions require a significant allocation of financial and organizational resources. For this reason, the most accurate early identification and preventive action (Balfanz et al., 2007) would be essential for any educational institution and country. One way to address this issue is through data mining (Pereira & Zambrano, 2017) and, in particular, prediction-based machine learning explainable models.

In the formal educational context, different studies were found about school dropout prediction. Alban and Mauricio (2019), using Artificial Neural Networks (ANN) with data from 2,670 university students over three years, found dropout prediction with 96.3% and 96.8% accuracy. Chung and Lee (2019) used Random Forest (RF) with data from 165,715 high school students with 95% accuracy. Pereira and Zambrano (2017), using Decision Tree (DT) obtained an accuracy above 80%. Fernández-García et al. (2021) used several models from the beginning of enrolment to the fourth semester identifying 91.5% who would drop out by the end of the fourth semester. Hutagaol and Suharjito (2019) used K-Nearest Neighbor (KNN), Naïve Bayes (NB) and DT reaching 79.12% accuracy in predicting college dropout. Kiss et al. (2019) using Gradient Boosted Tree (GB), XGB and ANN had an accuracy of up to 85.8%. Adnan et al. (2021), in a sample of 32,593 university students, achieved an accuracy of 91.9% with RF. The same algorithm was used by Berriri et al. (2021) on 154 students and by Uliyan et al. (2021) on 949 students obtaining, respectively, 80% and 93% accuracy. Ahmad and Shahzadi (2018), El Fouki et al. (2019) and He et al. (2020) used the ANN algorithm and, with samples of 300, 496 and 32,593 university students respectively, obtained accuracy rates of 93.2%, 92.54% and 91%. The DT algorithm was also used in the research of Barros et al. (2019) (n = 7,718), Freitas et al. (2020) (n = 1,549), Hamim et al. (2021) (n = 395), Nuankaew (2019) (n = 389) and Tan and Shao (2015) (n = 62,375) obtaining accuracy data of 95.39%, 99.34%, 94.07%, 87.21% and 94.63% respectively. Mourdi et al. (2020), with a sample of 49,551 university students achieved an accuracy of 89.8% with the Support Vector Machines (SVM) algorithm. Nabil et al. (2021) reached 76.2% with 4,266 students using Logistic Regression (LR) and Niyogisubizo et al. (2022) reached 92.89% with a smaller sample of 261 university students using a Stacking Ensemble (SE). Jadric et al. (2010) indicated that the most accurate models for predicting dropout were DT, LR and ANN. In a systematic review on the subject, Cardona et al. (2020) point out that the most commonly used AI models for predicting dropout in formal education are ANN, DT, LR, SVM, KNN, RF, NB; these same authors indicate that the most accurate models are ANN (71.5%-94%), DT (65.3%-81.3%), LR (50.1%-83%) and SVM (57.6%-86.4%). In summary, it is evident that in recent years the most frequent explanatory machine learning models that are being applied to the prediction of school dropout are ANN, RF, DT, KNN, NB, GB, SVM, LR and SE.

In this context, the central research questions are: are the predictions of AI models of school dropout accurate? and which are the most accurate AI models in predicting the dropout rate? To address these questions, the aim of this meta-analysis is to analyze the effectiveness of dropout prediction models. To do so, the empirical evidence from various international databases that employ explainable machine learning models in the prediction of school dropout in formal education is reviewed. The increase in the scientific literature published in recent years requires an updated assessment of the subject. The justification for this meta-analysis lies

in the fact that it is the first research which analyzes the effectiveness of the prediction of different AI models for school dropout and compares different models, revealing the most effective ones for this prediction.

However, a meta-analysis combines all the results of the collected AI studies on school dropout with the same focus, omitting those studies whose analyses are not based on concrete statistics. Moreover, the papers included in the meta-analysis imply a high heterogeneity in the way the original research is conducted, and this in practice is always a complex challenge for the meta-analysis itself which makes the answers to the questions approximately certain. In the present research, the meta-analysis is immersed in the variability of multiple factors such as the nature of the selected variables or factors used for training (academic, demographic and social), the type of parameters and/or hyperparameters used in each algorithm, the type of target group, the heterogeneity of the educational systems, the size and characteristics of the sample, the format and typology of the questions, as well as all those contextual variables that directly or indirectly are influencing modulating the results. In summary, these aspects highlight the difficulties of conducting a meta-analysis on the detection of school drop-out by means of AI. However, the scientific and academic field needs objective information that brings together all the existing research to know the accuracy of the models of AI in school dropout.

This is why the results of the meta-analysis will provide knowledge to help principals, managers and politicians make decisions to prevent school dropout and keep students in educational institutions. Specifically, they provide information on the most accurate explainable AI model they can use to predict the dropout rate of their educational institution. Thus, they can anticipate their students' school dropout and design the most appropriate measures to face that situation.

Methodology

In order to achieve the objectives of this study, a meta-analysis was carried out following the methodology proposed by Glass (1976), which allows an objective statistical analysis and greater generalization of the results when studies from different geographical areas with different backgrounds are used (Borenstein et al., 2009). The selection of articles for the meta-analysis was carried out following the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) methodology (Page et al., 2021). PRISMA contains 27 indicators that constitute a checklist and a flow chart with three phases (identification, screening and included) to facilitate the design and development of a protocol and critical evaluation (Page et al., 2021).

Search Criteria

This study uses both inclusion and exclusion eligibility criteria. Inclusion criteria are: (a) empirical research (articles, conferences) of AI in education, (b) focused on the prediction of school dropout, (c) in the context of formal education, (d) whose content is open access, and (e) in the social sciences field of knowledge.

The exclusion criteria are: (a) theoretical research (articles, book chapters and books), (b) in the context of non-formal or informal education, (c) not open access, (d) with a non-educational focus, and (e) the research does not provide the accuracy data necessary to perform the meta-analysis.

The reason for choosing these inclusion and exclusion criteria lies in the fact that this is a quantitative study, so we had to exclude theoretical research (exclusion criteria "a") and select exclusively empirical research (inclusion criteria "a"). On the other hand, although AI, after appropriate training, can be used to predict a wide range of situations, the present research has focused on the educational context (exclusion criteria "d"); in particular, the research has been directed at school dropout (inclusion criteria "b") and, therefore, within the formal education school context (inclusion criteria "c" and exclusion criteria "b"). Inclusion criteria "d" and exclusion criteria "c" and "e" ensured that the research content and, in particular, the effectiveness data of the trained models could be accessed in order to successfully carry out the meta-analytical study.

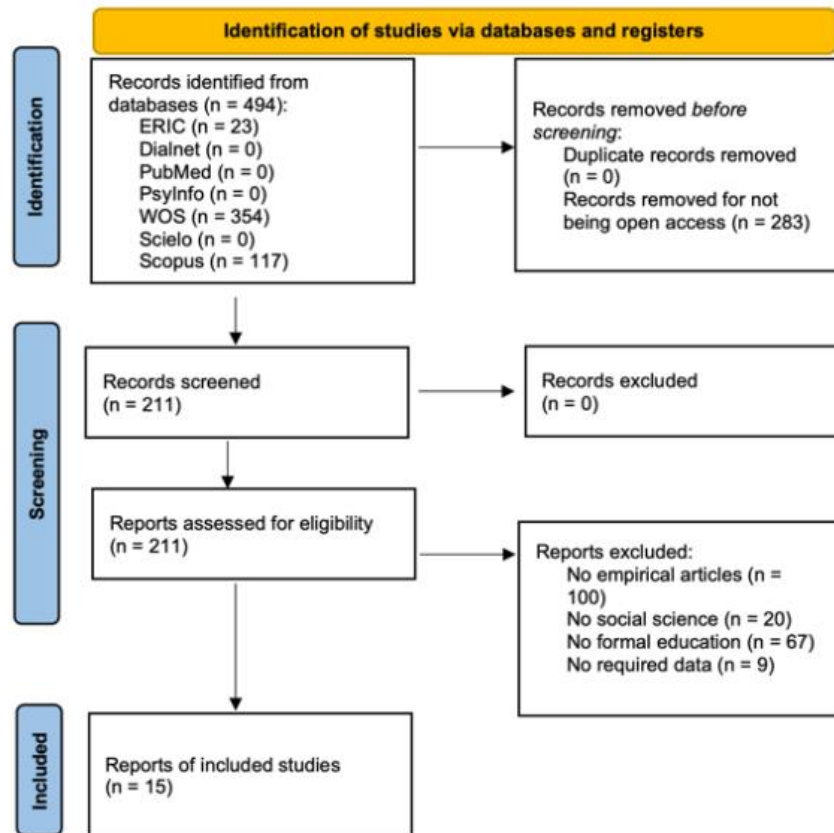
Search Strategies

The search, carried out in May 2022, was conducted in English and Spanish in the databases Web of Science (WOS), Scopus, PubMed, ERIC, PsycInfo, Dialnet and Scielo without limiting it to any time interval or language. In addition, the bibliographical references of the selected articles were reviewed to locate possible studies that met the inclusion criteria. The descriptors used in the title, abstract and keyword search were: "artificial intelligence", dropout and prediction in English and Spanish combined with Boolean operators AND and OR. No secondary sources/grey literature have been explored in this work. Thus, the resulting search string was the following:

*("artificial intelligence" OR "inteligencia artificial") AND (prediction OR predicción)
AND (abandono OR dropout)*

The search returned 494 publications in all databases, specifically 354 from WOS, 117 from Scopus and 23 from ERIC. Dialnet, PubMed, PsycInfo and Scielo returned no results. This total of 494 scientific productions is reduced to 15 articles that finally constitute the meta-analysis. The process followed from 494 is as follows: the articles were checked for full access (n = 211), they were articles (n = 111), the field of social sciences was selected (n = 91), they focused their research on formal education (n = 24) and they contained the data required for the meta-analysis (n = 15). No duplicate studies were detected in the sample. Fig. 1 shows the flow chart according to the PRISMA guidelines (Page et al., 2021) for the selection of publications included in the meta-analysis.

Figure 1
PRISMA Flow Chart



Codification Procedure

This study follows the guidelines of the systematic review manual (Higgins & Green, 2008) in which clear objectives, use of specific search terms and pre-defined eligibility criteria are indicated. PRISMA (Page et al., 2021) was used to search for empirical publications on the topic. The search process was conducted by two independent researchers with 100% agreement. A coding protocol is used to solve ambiguities, reflect on the proposals and resolve disagreements between the two researchers to 98%.

Table 1 systematizes a synthesis of the information of all the studies included in the work indicating authors/year, country of affiliation, country of data, sample size, educational level, type of data, AI models, metrics and effect size. This table is created by the researchers independently in Microsoft Excel to match the data.

For the moderator variables the coding was done as follows: secondary educational level (value = 0) or university educational level (value = 1); the number of variables used for the AI models, based on the average of the variables used in the 15 selected items, is less than fifteen (value = 0) or greater than fifteen (value = 1); the number of algorithms used in AI, based on the average of the algorithms used in the 15 selected articles, is less than five (value = 0) or greater than five (value = 1); the number of moments used is one (value = 0) or greater than

one (value = 1); and, the small sample size is less than one thousand (value = 0) or greater than one thousand (value = 1).

Effect Size and Statistical Analysis

This work had a sample of 15 papers using a continuous variable (accuracy) and proportions as effect size. Although other measures of AI effectiveness better compensate for the deviations produced by imbalanced samples, the authors of this article decided to analyse accuracy ($(True\ Positive + True\ Negative) / (True\ Positive + True\ Negative + False\ Negative + False\ Positive)$) because it was the only continuous variable shared by all the studies. In addition, calculating other measures better suited to imbalanced samples such as the F1-Score ($(Precision * Sensitivity) / (Precision + Sensitivity)$) proved impossible as several studies did not publish True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) data. All accuracy data were transformed to absolute frequencies and a random-effects model was used to calculate the effect size (Borenstein et al., 2009). This type of model allows for greater generalization of findings and an estimation of the effects of different sources of variation (moderating variables) (Borenstein et al., 2009). In this meta-analysis, the DerSimonian and Laird method (Egger et al., 2001) The raw proportion was calculated as well as for each proportion, the standard error and the confidence interval (CI) was of 95% and p.

To test the null hypothesis of homogeneity among the studies in the sample, Cochran's Q test was used, indicating with a significant value the presence of heterogeneity among the studies in the sample. To calculate the magnitude of this heterogeneity, variability I^2 was used to indicate the proportion of variability (proportion of observed variance that manifests the non-random difference between the effect sizes of the papers in the sample) (Higgins et al., 2003; Higgins et al., 2008). If I^2 reaches 25%, it is considered low level (differences due to random sampling and similar results between studies). However, if it reaches 50%, it is considered medium level and thus, 75% a high level (real differences due to different study designs) (Higgins et al., 2003). To analyze the effect of moderator variables on accuracy, a meta-regression analysis was used; for each moderator variable the estimation parameters, standard error, CI, Q and I^2 were calculated.

In the significant comparison between the different prediction models, the Chi-squared test recommended by Campbell (2007) and Richardson (2011) and the confidence interval according to the method of Altman et al. (2000) were used.

Although several international databases were researched, it is possible that relevant works may have been missed due to the search strategies used or any kind of bias. Two methods were used to assess publication bias: Rosenthal's fail-safe number and regression test. According to Borenstein et al. (2009) Rosenthal's test indicates the number of missing studies with null effects that would make the above correlations insignificant; when the values are large there is no bias. In regression test there is no bias when the regression is not significant.

The software used for data analysis in this work are Microsoft Excel, Jamovi version 1.6 and MedCalc (free statistical calculator from the internet).

Results

The results of the study are organized in three sections: the descriptive analysis of the data, the statistical analysis of the meta-analysis and the analysis of publication bias.

Description of the Studies Included

Fifteen papers were selected for the meta-analysis, all of them experimental studies. The publication range is between 2015 and 2022, most of them being published in 2021 (33%, 5/15). 6.6% were published in 2015, 6.6% in 2018, 20% in 2019, 26.6% in 2020 and 6.6% in 2022. Sample data for the research were obtained from 12 different countries. Data from two research studies were obtained from Brazil and the United Kingdom, while data from only one research work were obtained from China, France, Morocco, Pakistan, Portugal, Saudi Arabia, Slovakia, Spain, Thailand and the United States. In one of the studies (Nabil et al., 2021), the sample of the study belonged to 13 different countries.

The sample analyzed was produced by a total of 65 authors who, based on their countries of affiliation, came from China (20%, 13/65), Morocco (15.38%, 10/65), Pakistan (15.38%, 10/65), Brazil (13.85%, 9/65), Saudi Arabia (9.23%, 6/65), Spain (9.23%, 6/65), France (6.15%, 4/65), Egypt (4.62%, 3/65), Canada (1.54%, 1/65), USA (1.54%, 1/65), Rwanda (1.54%, 1/65) and Thailand (1.54%, 1/65).

This meta-analysis has a sample size of 199,015 participants, being the smallest size 154 students (Berriri et al., 2021) and the largest comprising a sample of 49,551 (Mourdi, et al., 2020). 46.6% (7/15) of the studies have a sample size of less than 1,000 participants, 26.6% (4/15) between 1,001-10,000 participants and 26.6% (4/15) larger than 10,001 participants. A summary of the selected studies is shown in Table 1.

Table 1

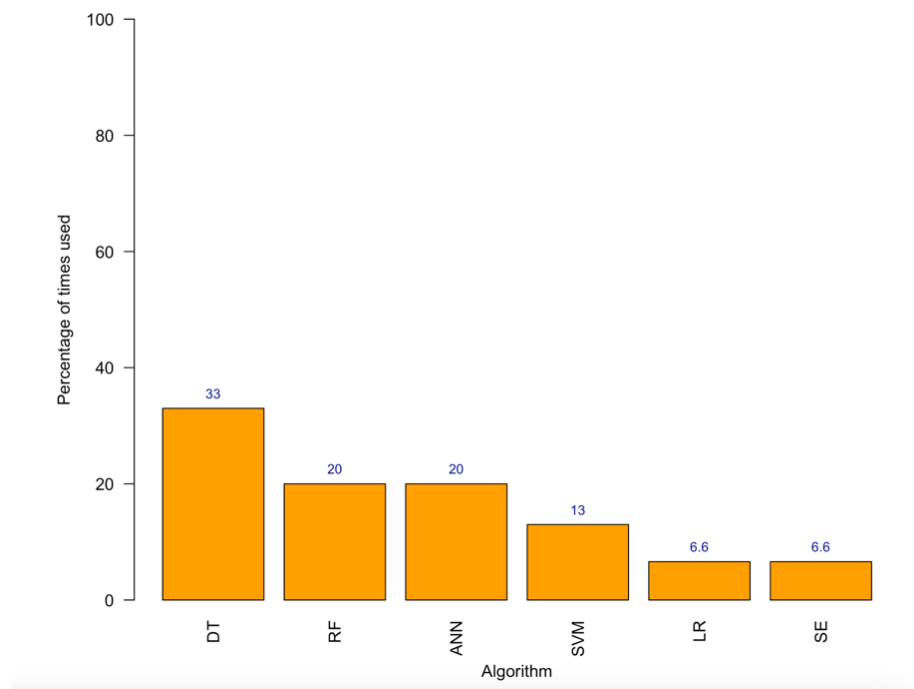
Results of the Empirical Studies Included in the Meta-Analysis

Authors/ year	Study		Participant Sample/ School level	Type of data	AI Model s	Metrics	Prop. Effect size
	Country of affiliatio n	Country of data					
Adnan et al., 2021	Pakistan	United Kingdom	29.953/ University	Academic and demographi c	RF, SVM, KNN, ET, AB, GB	Accuracy, Sensitivit y, Precision, F1-Score	0.92
Ahmad & Shahzadi, 2018	Pakistan	Pakistan	300/ University	Academic and demographi c	ANN	Accuracy	0.93

Authors/ year	Study		Participant s			AI Model s	Metrics	Prop. Effect size
	Country of affiliatio n	Country of data	Sample/ School level	Type of data				
Barros et al., 2019	Brazil	Brazil	7.718/ Secondary	Academic, demographi c and social		MLP, DT, BB	Accuracy, Sensitivit y, Precision, F1-Score	0.95
Berriri et al., 2021	France	France	154/ University	Academic		RF	Accuracy	0.86
El Fouki et al., 2019	Morocco	Morocco	496/ University	Academic		ANN, BN	Accuracy	0.93
Fernández- García et al., 2021	Spain	Spain	5.426/ University	Academic and demographi c		SVM, GB, RF, ENS	Accuracy, Sensitivit y, Precision	0.91
Freitas et al., 2020	Brazil, Canada, Saudi Arabia	Brazil	1.549/ University	Demographi c and social		DT, LR, SVM, KNN, MLP, DNN	Accuracy, Sensitivit y, Precision, F1-Score	0.99
Hamim et al., 2021	Morocco	Portugal	395/ Secondary	Academic, demographi c and social		DT, SVM, LR, KNN	Accuracy	0.94
He et al., 2020	China	United Kingdom	32.593/ University	Academic and demographi c		ANN	Accuracy, Sensitivit y, Precision	0.91
Mourdi et al., 2020	Morocco	USA	49.551/ University	Academic and demographi c		SVM, KNN, NB, LR, DT	Accuracy, Precision,	0.90
Nabil et al., 2021	Egypt	Kuwait, Jordan, Palestine, Iraq, Lebanon, Tunis, Saudi Arabia, Egypt, Syria, USA, Iran&Liby a, Morocco, Venezuela	4.266/ University	Academic		ANN, DT, LR, SVC, KNN	Accuracy, Sensitivit y, Precision, F1-Score	0.76

Authors/ year	Study		Participants			AI Models	Metrics	Prop. Effect size
	Country of affiliation	Country of data	Sample/ School level	Type of data				
Niyogisubizo et al., 2022	China, Rwanda	Slovakia	261/ University	Academic		ANN, RF, GB, SE	Accuracy, Sensitivity, Precision, F1-Score	0.93
Nuankaew, 2019	Thailand	Thailand	389/ University	Academic		DT	Accuracy, Sensitivity, Precision	0.87
Tan & Shao, 2015	China	China	62.375/ University	Academic and demographic		ANN, DT, BN	Accuracy, Sensitivity, Precision, F1-Score	0.95
Uliyan et al., 2021	Saudi Arabia, Egypt, USA	Saudi Arabia	949/ University	Academic and demographic		LR, DT, RF, NB, SVM, ANN	Accuracy, Sensitivity, Precision, F1-Score	0.93

In the analyzed papers, 33% (5/15) used the DT model for prediction, 20% (3/15) RF, 20% (3/15) ANN, 13% (2/15) SVM, 6.6% (1/15) LR and 6.6% (1/15) SE. This distribution of algorithms used in the sample analyzed is graphically represented in Figure 2.

Figure 2*AI Algorithms Used in the Sample*

In 11 of the 15 studies analyzed, the algorithms were experimented in isolation, without using any procedure that combined and decided between the results of several algorithms (ensembles). In those where some form of ensemble was used, it was treated as another algorithm that was compared with the rest. For example, in the study by Barros et al. (2019) a Balanced Bagging technique with different DTs is used. In the paper by Fernández-García et al. (2021), one of the algorithms used for comparison is an ensemble that combines the results of a GB, a RF and a SVM. Similarly, a Stacking Ensemble is considered as a further option for comparison in the paper of Niyogisubizo et al. (2022). However, the study by Mourdi et al. (2020) is the only one that uses an ensemble in which the results of the SVM, KNN, DT, NB and LR algorithms are voted, but since the study provides the data independently per algorithm, these have been analyzed in isolation for the present research.

According to the moderating variables for the meta-analysis, the predominant educational level is higher education with 86.6% (13/15) of the studies and, to a lesser extent, secondary education with 13% (2/15) of the studies. 53% (8/15) use more than fifteen variables and 46.6% (7/15) less than fifteen variables. 73.3% (11/15) use a single data collection moment and 26.7% (4/15) use more than one. 53% (8/15) use five or more algorithms and 46.6% (7/15) less than five algorithms.

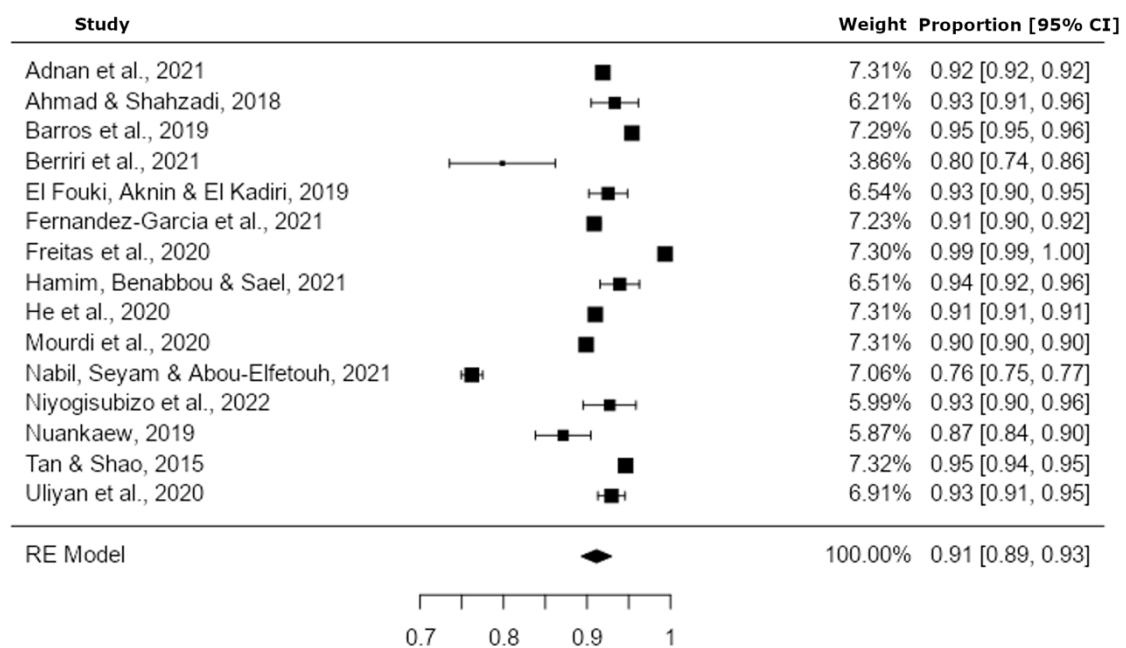
Results of the Random-Effects Model Meta-Analysis

In the meta-analysis, a random-effects model with the accuracy of all studies was used. The accuracy shows a mean proportion effect size of 91% (95% CI = 89-93%). The forest plot is shown in Figure 3. The variability between the different samples was significant ($Q =$

2843.805, $p < 0.001$; $I^2 = 99.51\%$). The type of model used ($p < 0.001$) had a significant contribution to heterogeneity, indicating that the dropout effectiveness ratio decreases using RF, ANN, SVM, LR and SE models. The other moderating variables used in this paper did not have a significant contribution to heterogeneity: educational stage ($p = 0.158$), number of algorithms used ($p = 0.836$), number of moments used ($p = 0.682$), number of variables ($p = 0.468$), sample size ($p = 0.849$). This high variability is explained by other factors not considered in this study.

Figure 3

Forest Plot Dropout Accuracy



The study tested the efficacy of the different dropout models used. The DT model has a mean proportion effect size of 95.3% (95% CI = 93-98%) with significant variability ($Q = 460.722$, $p < 0.001$; $I^2 = 99.13\%$). The RF model has a mean proportion effect size of 90.6% (95% CI = 88-93%) with significant variability ($Q = 15.423$, $p < 0.001$; $I^2 = 87.03\%$). The ANN model has a mean proportion effect size of 91.8% (95% CI = 90-93%) with significant variability ($Q = 4.203$, $p < 0.001$; $I^2 = 52.42\%$). The remaining SVM, LR and SE models cannot be calculated due to a sample size of less than 3 studies.

Significant comparison between the three models shows that the DT model is significantly better than RF ($X^2 = 874.511$; difference = 4.7%; CI (4.3554-5.0508%); $p < 0.0001$) and ANN ($X^2 = 514.606$; difference = 3.5%; CI (3.1720-3.8342%); $p < 0.0001$). Comparing the RF and ANN models, ANN predicts dropout significantly better ($X^2 = 30.226$; difference = 1.2%; CI (0.7723-1.6279%); $p < 0.0001$).

In the DT model, which is the most effective at predicting dropout, possible moderating variables have been analyzed, but none of them contribute significantly to explain the heterogeneity: educational stage ($p = 0.742$), number of algorithms ($p = 0.309$) and number of

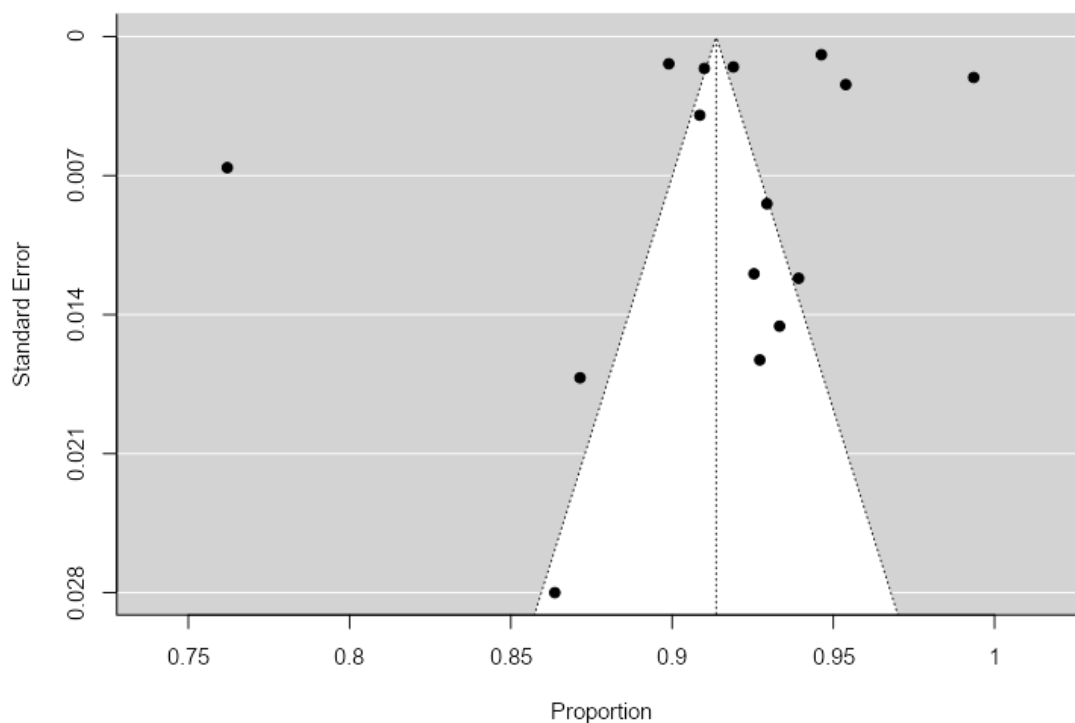
variables ($p = 0.334$). This high variability is explained by other factors not considered in this study. In the RF and ANN models, as there are fewer than 3 studies, the analysis of moderating variables cannot be performed.

Analysis of Publication Bias

Rosenthal's fail-safe number (Fail-safe $N = 7837746.00$, $p < 0.001$) and regression test ($Z = -1.321$, $p = 0.187$) were used for this analysis. Neither procedure showed any evidence of publication bias in the studies. Figure 4 shows the funnel plot, each point representing a study included in the meta-analysis.

Figure 4

Funnel plot dropout accuracy



For studies using DT (Fail-safe $N = 1670610.000$, $p < 0.001$) and regression test ($Z = -0.235$, $p = 0.814$) there is no evidence of publication bias in the studies.

Discussion

In recent years, several AI models have been applied to predict the dropout rate in different educational levels (mostly higher education) of formal education. So far, there is not any meta-analysis that systematizes all empirical research on the subject and compares the AI models used. Therefore, finding out whether AI models are being accurate in different countries and identifying the best predictive model of early school dropout is an internationally relevant issue for all governments and educational leaders.

This research aimed to analyze the effectiveness of dropout prediction models in formal education. Specifically, this meta-analysis found that the most commonly used AI machine learning models in scientific research are highly accurate. Moreover, the most accurate and effective AI model for dropout prediction is DT followed by ANN and RF.

Are the Predictions of AI Models of School Dropout Accurate?

The results of this paper show that the explanatory AI models used in the studies perform well in predicting school dropout, although this statement should be taken with caution since the only continuous variable shared by all the studies analyzed was accuracy, whose limitations are described in studies such as that of Sokolova et al. (2006), and that, as previously described, most of the studies analyzed did not provide the data to calculate other more appropriate measures. However, this finding is in line with Pereira and Zambrano (2017) who indicate that the analyzed models offer high confidence predictions in school dropout (Delen, 2011; Dissanayake, Robinson & Al-Azzam, 2016). They also indicate that they are efficient methods for predicting the dropout rate in educational institutions. Studies such as Martinho et al.'s (2013) and Kostopoulos et al.'s (2017) show a high success rate in detecting dropout at a very soon stage, exclusively with data extracted from the enrolment process or at an early point in the course schedule. This finding is internationally relevant because this dropout indicator represents a measure of the success of the educational institution and of the students.

Five factors were quantitatively assessed in the analysis of moderating variables. Only one of them, the type of AI model, was relevant as far as it indicated that the dropout prediction ratio improves using DT compared to other unspecified models (RF, ANN, SVM, LR and SE). According to Adejo and Connolly (2018) the advantage of DT over other models is based on the flexibility to model non-linearity and computational speed; it also offers a more visual and easier to understand the results. For Dissanayake et al. (2016) the DT model performs better when using original data rather than principal components and for Cardona et al. (2019) the DT is an effective AI model for predicting dropout.

The remaining moderating variables were not significant. The educational stage (high school vs. university), the number of variables used for the AI models, the number of algorithms used in AI, the number of data collection moments used, and the sample size did not explain the variability in the prediction of school dropout. Therefore, there are other factors not captured in this meta-analysis that moderate and explain the heterogeneity in dropout prediction. This implies that the results of this paper should be interpreted with caution and in the context of the sample used.

Which are the Most Accurate IA Models when Predicting the Dropout Rate?

When comparing the different AI models, the DT model, when predicting dropout, has a significantly higher proportion (95.3%) than the other models (RF, ANN, SVM, LR and SE) used in the study of the subject. The findings of the meta-analysis indicate that the DT model is the most accurate predictor of dropout in formal education. This result is in line with the research conducted by other authors such as Sreenivasa et al. (2018) who found that the DT

model was the most accurate model compared to other models such as RF, NB and J48 when analyzing first undergraduate students; Freitas et al. (2020) demonstrated that DT achieves the best results in accuracy, recall, precision and F1 Score when analyzing undergraduate students using socioeconomic data and different algorithm (DT, LR, SVM, KNN, MLP, DNN). For Hamim et al. (2021) DT provided the most accurate performance in the context of traditional education. On the other hand, the results of this work are in contrary to those obtained by Delen (2011) who, with a sample of 25,224 first-year university students and using the ANN, DT and RL models, attained that ANN is the model that performed best when predicting the dropout rate, indicating that the economic factor and past and present academic performance are the most relevant parameters. Oztekin (2016) used data from university students in a time span of four years, and compared three models: ANN, DT and SVM to predict school completion. He found out that the most accurate model was SVM.

As for the comparison between RF and ANN models, the most accurate prediction was made by ANN. These results are in line with several studies. The research by Mduma and Machuve (2021) obtained higher results in the case of ANN (83.6%), compared to RF (77.6%). The study carried out by Sani et al. (2020) showed that there was hardly any significant difference in accuracy values between RF and ANN algorithms (95.93% and 95.86% respectively).

Conclusion

This meta-analysis identifies the most accurate machine learning (ML) models for predicting dropout in formal education. This meta-analysis found that the AI models currently being used in dropout rate prediction are accurate and perform well in predicting dropout rates. Moreover, the prediction accuracy ratio is higher using the DT model versus other models (RF, ANN, SVM, LR and SE) and using DT versus ANN and RF, with the use of factors of the type academic (academic achievement, previous academic achievement, e-learning performance, educational background), social (economic, parental education, living place) and demographic (gender, age, nationality, academic level, marital status). If we look at the characteristics of these factors, which are heterogeneous and have different scales of measurement, we can find one of the reasons why the DT algorithm has proved to be more effective than others in predicting school dropout. In this sense, Pal and Mather (2003) state that DT can handle data measured with different scales. Moreover, it does not behave as a black box, as in other algorithms, but rather an analyst can examine the decision tree after the learning process, where each of the previously described factors (academic, social or demographic) can constitute a node of the tree that leads to the final prediction of dropout or not dropout.

The practical implications are that educational systems can implement appropriate measures at the optimal time for students to continue studying and finally, graduate. In the field of management by principals and managers of educational institutions, the main contribution is to guide and improve the decision-making procedure by controlling the supply of updated information and assessment of the documentation of the centres. In this sense, it can be a tool to reduce risk and uncertainty when making an investment decision on school enrolment and prognosis of vulnerable dropout conditions. For example, making a decision on the promotion or repetition of students who have failed a grade and proposing other alternatives for failing

students, such as re-examination or repeating the failed grade. At the same time, this field of management entails the design of educational strategies and the implementation of academic actions such as, on the one hand, early warning systems can be set up to detect students who may be at risk during their first year at university or high school. Tsao et al. (2017) designed an early warning system using DT and a heuristic model that is effective for their analyzed educational institution. On the other hand, individualized intervention measures can be designed through counselling for each student by providing appropriate actions to keep them in the educational system, improve their academic performance and increase the success rate of the educational institution, e.g. motivational and attention-grabbing strategies to keep students focused on their studies. Heublein (2014), in his review, proposes to extend the support offered to students during the start of studies, better information and more flexible curricula in higher education.

In general, AI models will be applied to the educational context more and more frequently, so leaders, teachers and researchers should be constantly updated to provide the most appropriate actions for each student, avoiding school dropout and optimizing their academic and personal success. According to the results of this meta-analysis for dropout prediction the use of DT is pertinent in empirical investigations when using DT with optimal hyperparameters as identified or recommended in previous studies (Gomes et al., 2018). In this way they can obtain maximum information by improving the model architecture, because the adjustment of hyperparameter values is an alternative to achieve optimal predictive performance, i.e. DT will make better predictions on invisible data, which it cannot learn from, helping to prevent overfitting.

Limitations

To interpret the results of this work, it should be noted that this meta-analysis focuses on 15 empirical articles that met the inclusion criteria. Therefore, the sample was limited to 15 independent primary studies. The results cannot be generalized to other research that does not indicate accuracy. This limitation of the accuracy data is relevant because some articles do not provide it and use F1-Score instead because it has fewer problems with imbalanced data. The issue of imbalanced data is a major challenge in sample classification problems by AI algorithms, because the effectiveness of the algorithm increases for the category with higher frequency of occurrence, while it decreases for the category with lower frequency of occurrence in AI training processes (Jiang et al., 2012; Su et al., 2006). This problem has been addressed in different ways in some of the papers studied. For example, the research by Barros et al. (2019) describes the downsampling or upsampling techniques used to mitigate the imbalance problem. The paper by Berriri et al. (2021) uses two ways to balance the sample prior to AI training: on the one hand, grouping the more numerous classes of students into subclasses with a better balance and, on the other hand, adjusting the parameters of the RF algorithm so that the classes with fewer samples have a higher weight, adjusting their weight inversely proportional to their frequency of appearance in the data. The study by Nabil et al.

(2021) compares different techniques for sample balancing, such as SMOTE, ADASYN, ROS and SMOTE-ENN.

A limitation also related to the primary school sample is the educational level, mostly in higher education. Research focuses more on university students and to a lesser extent on secondary or other lower educational stages. The main reason is the fact that the highest dropout rates are found in higher education; the OECD (2019) estimates a 30% dropout rate at university in all OECD member countries. This implies an economic and social cost for universities and governments.

In the search descriptors, AI is included but machine learning and supervised/unsupervised learning are not included. AI is used in this research as it is considered to be a more generic descriptor and could provide more research results. However, this is possibly a limitation and some studies may have been overlooked, so it may be useful to take this into account in future meta-analyses on the subject.

One aspect that depends on the primary studies used in the meta-analysis is the higher representation of certain Asian countries, since 46.15% of the authors, according to their affiliation data, belonged to Asian countries, followed by 21.54% of authors from countries on the African continent. This may be related to the funding received for this type of AI-based studies or the interests of certain countries in developing this type of technology. This may bias the results and should be taken into account in the interpretation of this meta-analysis.

Another factor to take into account, is the high heterogeneity among the studies, which is not explained by the moderating variables analyzed in this paper, except for the type of model. This implies interpreting the results in the context of the sample used and conducting an in-depth analysis.

The last important point refers to the variables that each study uses to construct the AI model of dropout prediction. Some studies focus more on academic factors (Berriri et al., 2021; El Fouki et al., 2019; Nabil et al., 2021; Niyogisubizo et al., 2022; Nuankaew, 2019), others combine academic and demographic data (Adnan et al., 2021; Ahmad & Shahzadi, 2018; Fernández-García et al., 2021; He et al., 2020; Mourdi et al., 2020; Tan & Shao, 2015; Uliyan et al., 2021), whereas a final, smaller group analyses academic, demographic and social data (Barros et al., 2019; Hamim et al., 2021). As a particular case that has not been included in any of the three previous groups, the study by Freitas et al. (2020) used demographic and social data, but not academic data. This heterogeneity in the use of different factors in the articles analyzed means that the results are not entirely homogeneous because they do not originate from the same type of data, i.e. findings from academic and demographic data alone are mixed with demographic and social and academic and social data. In this sense, Tsao et al. (2017) concluded in their study that the selection of variables to generate the data for the AI model considerably influences the performance of these prediction models. At the same time, we are aware of the relevance of generating an AI model from as much data as possible from different factors. In fact, it can be seen that dropout prediction depends not only on socioeconomic factors but also on the presence of other factors such as academic performance, personal circumstances or lack of motivation. For this reason, studies on AI models for dropout prediction could possibly try to obtain the largest volume of data from different factors and categorise the results by independent and combined factors, so that data of the same nature can be compared (even knowing that, within the same factor, there is a diversity of data). In the

coming years, when there are more articles on the subject, meta-analyses can be carried out that evaluate the same data, for example, demographic and social data alone, or academic and social data alone, and can be compared to see which factors and categories within these factors are the most effective for dropout prediction alarm systems.

Prospective

Future lines of research need to be directed towards the application of DT as the most accurate model and the combination of DT plus other models that are the most accurate ones in each educational institution; personalization in schools and universities is essential. At the same time, more research is needed in other educational stages different from higher education to increase the transition of students to higher education.

For future work, it is recommended to incorporate grey literature, expand the search databases and extend research to recommendation systems and adaptive learning. Knowing the prediction data is paramount and requires later recommendation actions for each student and institutions to improve the quality of education.

References

- Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, 10(1), 61-75. <https://doi.org/10.1108/JARHE-09-2017-0113>
- Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., et al. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *Ieee Access*, 9, 7519-7539. Doi: [10.1109/ACCESS.2021.3049446](https://doi.org/10.1109/ACCESS.2021.3049446)
- Ahmad, Z., & Shahzadi, E. (2018). Prediction of Students' Academic Performance Using Artificial Neural Network. *Bulletin of Education and Research*, 40(3), 157-164.
- Alban, M., & Mauricio, D. (2019). Neural networks to predict dropout at the universities. *Int. J. Mac. Learn. Comput.*, 9, 149-153. Doi: [10.18178/ijmlc.2019.9.2.779](https://doi.org/10.18178/ijmlc.2019.9.2.779)
- Aldabas-Rubira, E. (2002). Introducción al reconocimiento de patrones mediante redes neuronales. *IX Jornades de Conferències d'Enginyeria Electrònica del Campus de Terrassa, Terrassa, España, del 9 al 16 de Diciembre del 2002*.
- All, E. (2019). *Education at a Glance 2019 OCDE Indicators*. OECD.
- Altman, D.G., Machin, D., Bryant, T.N., & Gardner, M.J. (2000). *Statistics with confidence*. BMJ Books.
- Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Nesta. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- Balfanz, R., Herzog, L. & MacIver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235. <https://doi.org/10.1080/00461520701621079>
- Barros, T. M., Souza Neto, P. A., Silva, I., & Guedes, L. A. (2019). Predictive models for imbalanced data: a school dropout perspective. *Education Sciences*, 9(4), 275. <https://doi.org/10.3390/educsci9040275>
- Berriri, M., Djema, S., Rey, G., & Dartigues-Pallez, C. (2021). Multi-class assessment based on random forests. *Education Sciences*, 11(3), 92. <https://doi.org/10.3390/educsci11030092>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons, Ltd.
- Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine* 26, 3661-3675. DOI: [10.1002/sim.2832](https://doi.org/10.1002/sim.2832)
- Cardona, T., Cudney, E.A., & Snyder, J. (2019). Predicting degree completion through data mining. *126TH Annual Conference & Exposition*, del 15-19 June 2019. doi:[10.18260/1-2-33183](https://doi.org/10.18260/1-2-33183)
- Cardona, T., Cudney, E.A., Hoerl, R. & Snyder, J. (2020). Data mining and machine learning retention models in higher education. *Journal of College Studente Retention: Reseachr, Theory & Practicec*, 0(0), 1-25. doi:[10.1177/1521025120964920](https://doi.org/10.1177/1521025120964920)
- Chung, J., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Child. Youth Serv. Rev.*, 96, 346-353. <https://doi.org/10.1016/j.chilyouth.2018.11.030>

- Day, T., Chang, I. C., Chung, C. K. L., Doolittle, W. E., Housel, J., & McDaniel, P. N. (2021). The immediate impact of COVID-19 on postsecondary teaching and learning. *Professional Geographer*, 73(1), 1-13. doi:[10.1080/00330124.2020.1823864](https://doi.org/10.1080/00330124.2020.1823864)
- Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory and Practice*, 13(1), 17–35. <https://doi.org/10.2190/CS.13.1.b>.
- Dissanayake, H., Robinson, D., & Al-Azzam, O. (2016). Predictive Modeling for Student Retention at St. Cloud State University. In *Proceedings of the International Conference on Data Mining (DMIN)* (p. 215). WorldComp.
- Dobrev, D. (2012). A definition of artificial intelligence. *arXiv preprint:1210.1568*. <https://doi.org/10.48550/arXiv.1210.1568>
- Du, X., Yang, J., & Hung, J-L. (2020). An integrated framework based on latent variational autoencoder for providing early warning of at-risk students. *IEEE Access*, 8, 10110-10122. Doi: [10.1109/ACCESS.2020.2964845](https://doi.org/10.1109/ACCESS.2020.2964845)
- Educause (2019). *Horizon report: 2019 higher education edition*. <https://library.educause.edu/media/files/library/2019/4/2019horizonreport.pdf>
- Egger, M., Smith, D., & Altmand, D. G. (2001). *Systematic Reviews in Health Care: Meta-Analysis in Context*. BMJ Publishing Group.
- El Fouki, M., & Aknin, N. (2019). Multidimensional Approach Based on Deep Learning to Improve the Prediction Performance of DNN Models. *International Journal of Emerging Technologies in Learning*, 14(2). <https://doi.org/10.3991/ijet.v14i02.8873>
- Fernández-García, A., Preciado, J., Melchor, F., Rodriguez-Echevarria, R., Conejero, J., & Sanchez-Figueroa, F. (2021). A real-life machine learning experience for predicting university dropout at different stages using academic data. *IEEC Access*, 9, 133076-133090. <https://doi.org/10.1109/ACCESS.2021.3115851>
- Freitas, F. A. D. S., Vasconcelos, F. F., Peixoto, S. A., Hassan, M. M., Dewan, M. A. A., Albuquerque, V. H. C. D., & Filho, P. P. R. (2020). IoT system for school dropout prediction using machine learning techniques based on socioeconomic data. *Electronics*, 9(10), 1613. <https://doi.org/10.3390/electronics9101613>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8. <https://doi.org/10.3102/0013189X005010003>
- Gomes Mantovani, R., Horváth, T., Rossi, A.L.D., Cerri, R., Barbon Junior, S., Vanschoren, J., & Pone de Leon Ferreira de Carvalho, A.C. (2018). Better Trees: An empirical study on hyperparameter tuning of classification decision tree induction algorithms. *arXiv*, [1812.02207](https://arxiv.org/abs/1812.02207)
- González, J., & Balaguer, A. (2007). Revisión sistemática y metanálisis (I): conceptos básicos. *Evid Pediatr.*, 3, 107. Doi: [vol3/2007_numero_4/2007_vol3_numero4.23.htm](https://doi.org/10.2190/CS.13.1.b)
- Hamim, T., Benabbou, F., & Sael, N. (2021). Survey of machine learning techniques for student profile modeling. *International Journal of Emerging Technologies in Learning (iJET)*, 16(4), 136-151. <https://doi.org/10.3991/ijet.v16i04.18643>
- He, Y., Chen, R., Li, X., Hao, C., Liu, S., Zhang, G., & Jiang, B. (2020). Online at-risk student identification using RNN-GRU joint neural networks. *Information*, 11(10), 474. <https://doi.org/10.3390/info11100474>

- Heublein, U. (2014). Student Drop-out from German Higher Education Institutions. *Eur. J. Educ.*, 49, 497-513. <https://doi.org/10.1111/ejed.12097>
- Higgins, J. P., & Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327, 327–557. doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)
- Hutagaol, N., & Suharjito, S. (2019). Predictive modelling of student dropout using ensemble classifier method in higher education. *Adv. Sci. Technol. Eng. Syst. J.*, 4, 206-2011. Doi: [10.25046/aj040425](https://doi.org/10.25046/aj040425)
- Jadric, M., Garaca, Z. & Cukusic, M. (2010). Student dropout analysis with application of data mining methods. *Management: Journal of Contemporary Management Issues*, 15(1), 31-46.
- Jiang, S., Pang, G., Wu, M., & Kuang, L. (2012). An improved K-nearest-neighbor algorithm for text categorization. *Expert Systems with Applications*, 39(1), 1503-1509. [10.1016/j.eswa.2011.08.040](https://doi.org/10.1016/j.eswa.2011.08.040)
- Kearney, C. A. (2008a). An interdisciplinary model of school absenteeism in youth to inform professional practice and public policy. *Educational Psychology Review*, 20, 257–282. <https://doi.org/10.1007/s10648-008-9078-3>.
- Kim, D., & Kim, S. (2018). Sustainable education: Analyzing the determinants of university student dropout by nonlinear panel data models. *Sustainability*, 10(4), 954. <https://doi.org/10.3390/su10040954>
- Kiss, B., Nay, M., Molontany, E., & Csabay, B. (2019). Predicting dropout using high school and first-semester academic achievement measures. In *Proceedings of the 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, Starý Smokovec, Slovakia, 21–22 November 2019.
- Kostopoulos, G., Kotsiantis, S., Ragos, O., & Grapsa, T. N. (2017). Early dropout prediction in distance higher education using active learning. In *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1-6). IEEE.
- Mahoney, J. L. (2018). School dropout. In M.H. Bomstein & M.E. Arterberry, K.L. Fingeman & E. Landsford (Eds.), *The SAGE encyclopedia of lifespan human development* (pp.1889-1891). SAGE.
- Martinho, V. R. D. C., Nunes, C., & Minussi, C. R. (2013). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence* (pp. 159-166). Ieee.
- Mduma, N., & Machuve, D. (2021). Machine Learning Model for Predicting Student Dropout: A Case of Tanzania, Kenya and Uganda. In *2021 IEEE AFRICON* (pp. 1-6). IEEE.
- MedCalc Software Ltd. Comparison of proportions calculator. https://www.medcalc.org/calc/comparison_of_proportions.php (Version 20.110; accessed June 9, 2022).
- Mira, J., Delgado, A.E., Boticario, J.G., & Díez, F.J. (2003). *Aspectos básicos de la inteligencia artificial*. Sanz y Torres.
- Moschovakis, Y. N. (2001). What is an algorithm?. In *Mathematics unlimited—2001 and beyond* (pp. 919-936). Springer, Berlin, Heidelberg.

- Mourdi, Y., Sadgal, M., Fathi, W. B., & El Kabtane, H. (2020). A machine learning based approach to enhance MOOC users' classification. *Turkish Online Journal of Distance Education*, 21(2), 47-68.
- Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access*, 9, 140731-140746. [10.1109/ACCESS.2021.3119596](https://doi.org/10.1109/ACCESS.2021.3119596)
- 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, 100066. <https://doi.org/10.1016/j.caeai.2022.100066>
- Nuankaew, P. (2019). Dropout situation of business computer students, University of Phayao. *International Journal of Emerging Technologies in Learning (iJET)*, 14(19), 115-131. <https://doi.org/10.3991/ijet.v14i19.11177>
- OECD (2019). *Education at a Glance 2019: OECD indicators*. OECD Publishing, Paris
- Oztekin, A. (2016). A hybrid data analytic approach to predict college graduation status and its determinative factors. *Industrial Management & Data Systems*, 116(8) 1678-1699. <https://doi.org/10.1108/IMDS-09-2015-0363>
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Hoffmann, T.C., Mulrow, C.D. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71). Doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)
- Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, 86(4), 554-565.
- Pereira, R., & Zambrano, J. (2017). Application of decision trees for detection of student dropout profiles. *Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Cancun, Mexico, 18–21 December 2017.
- Popenici, S., & Kerr, S. (2017). Exploring the analysis artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(22). <https://doi.org/10.1186/s41039-017-0062-8>
- Reuge, N., Jenkins, R., Brossard, M., Soobrayan, B., Mizunoya, S., Ackers, J., et al. (2021). Education response to COVID 19 pandemic, a special issue proposed by UNICEF: Editorial review. *International Journal of Educational Development*, 87, 102485. Doi: [10.1016/j.ijedudev.2021.102485](https://doi.org/10.1016/j.ijedudev.2021.102485)
- Richardson, J.T.E. (2011). The analysis of 2 x 2 contingency tables – Yet again. *Statistics in Medicine* 30, 890. Doi: [10.1002/sim.4116](https://doi.org/10.1002/sim.4116)
- Sani, N. S., Nafuri, A. F. M., Othman, Z. A., Nazri, M. Z. A., & Mohamad, K. N. (2020). Drop-out prediction in higher education among B40 students. *International Journal of Advanced Computer Science and Applications*, 11(11). [10.14569/IJACSA.2020.0111169](https://doi.org/10.14569/IJACSA.2020.0111169)
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence* (pp. 1015-1021). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Sreenivasa Rao, K., Swapna, N., & Praveen Kumar, P. (2018). Educational data mining for student placement prediction using machine learning algorithms. *International Journal of*

- Engineering and Technology (UAE)*, 7(1.2), 43–46.
<https://doi.org/10.14419/ijet.v7i1.2.8988>
- Su, J.S., Zhang, B.F., & Xu, X. (2006). Advances in machine learning based text categorization. *Ruan Jian Xue Bao (Journal of Software)*, 17(9), 1848-1859.
- Tan, M., & Shao, P. (2015). Prediction of student dropout in e-Learning program through the use of machine learning method. *International journal of emerging technologies in learning*, 10(1). <https://doi.org/10.3991/ijet.v10i1.4189>
- Tsao, N. L., Kuo, C. H., Guo, T. L., & Sun, T. J. (2017, July). Data Consideration for At-Risk Students Early Alert. In *Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress on* (pp. 208-211). IEEE.
- Uliyan, D., Aljaloud, A. S., Alkhalil, A., Al Amer, H. S., Mohamed, M. A. E. A., & Alogali, A. F. M. (2021). Deep learning model to predict students retention using BLSTM and CRF. *IEEE Access*, 9, 135550-135558. Doi:[10.1109/ACCESS.2021.3117117](https://doi.org/10.1109/ACCESS.2021.3117117)
- UNICEF (2017). Improving education participation. Policy and Practice Pointers for Enrolling All Children and Adolescents in School and Preventing Dropout. *UNICEF Series on Education Participation and Dropout Prevention*, 2.
- UNICEF (2018). Early warning systems for students at risk of dropping out. *UNICEF Series on Education Participation and Dropout Prevention*, 2.
- Wilcoxson, L., Cotter, J., & Jon, S. (2011). Beyond the first-year experience: The impact on attrition of student experiences throughout undergraduate degree studies in six diverse universities. *Stud. High. Educ.*, 36, 331-352. <https://doi.org/10.1080/03075070903581533>
- Xiao, J., Wang, M. Jiang, B., & Li, J. (2018). A personalized recommendation system with combinational algorithm for online learning. *J. Ambient Intell. Humanized Comput.*, 9(3), 667-677. [10.1007/s12652-017-0466-8](https://doi.org/10.1007/s12652-017-0466-8)