

Mitigating Academic Institution Dropout Rates with Predictive Analytics Algorithms

Bongs Lainjo

Cybermatic International, Canada

bsuiru@icloud.com

Article Information:

Received: September 28, 2022, Accepted: December 15, 2022, Published: January 02, 2023

Abstract

Several research findings show that student dropout rates differ between rich, low, and middle-income nations. With the advancement of machine learning, numerous academic institutions have begun to implement predictive analytics algorithms to address this issue. With great focus on North America, the research is aimed to comprehensively review PAAs Strategies and demonstrate how they have been properly implemented in maximizing retention rates. To review the existing literature, the study used a quantitative research methodology and a thematic study design. The result shows that learning institutions used a combination of data mining classifiers such as k-Nearest Neighbor, Neural Networks, Decision Tree, and Naive Bayes to categorize student dropout rates and optimize retention rates. The PAAs approach has been utilized by various educational institutions across North America, applying multiple predicting algorithms to ensure efficiency.

Keywords: Predictive Analytics Algorithms, Mitigating Dropout Rates, Retention Rates, Classifiers, Learning Institutions, North America

1. Introduction

Numerous academic institutions worldwide are struggling to reduce student drop-out rates. Findings suggest that significant student drop - out rates emerge in both developed and developing nations (Baker et al., 2020; Blazer & Hernandez, 2018). In recent decades, there has been much interest in using machine learning through predictive analytical algorithms (PAAs) to address this issue. Numerous scholars have published much literature on predicting school drop-outs (Marôco et al., 2020; Kemper et al., 2020). They have adopted different machine learning approaches to construct PAAs, including decision trees, artificial neural

networks, k-Nearest Neighbor, and Naive Bayes. Therefore, it is essential to explore how institutions utilize PAAs strategies in mitigating school drop-outs.

1.1 General Objective

To conduct a comprehensive review of Predictive Analytics Algorithms (PAAs) Strategies and demonstrate how they have effectively optimized retention rates with a case study focusing on North America.

1.2 Background

Teachers, school officials, communities, and the government have worked to lower drop-out rates for decades, yet it remains a concern in schools today. Leaving school is a critical educational topic in North America and a global social-economic problem. Before recent technological and sociological advances, it was increasingly difficult for people without a bachelor's degree to find work (Rastrollo-Guerrero et al., 2020). Aside from the increased risk of underemployment or unemployment, individuals who leave school are more susceptible to psychological conditions such as depression, joining a gang, or engaging in other illegal activities. In addition, there is evidence of the difference in wages between graduate and non-graduate students (Dvouletý, 2018) – an issue that should concern those vulnerable students. Most researchers usually adopt data mining and analytics methods to tackle issues in education. The approach ranges from determining which mathematical problems depict the same analytical capacity to detecting various students losing interest throughout the learning process and effective mechanisms to implement sustainable intervention (Coussement et al., 2020; Evans et al., 2020; Gkontzis et al., 2019). PAAs, whether built in-house or procured from third-party providers or the government, have demonstrated a green light in mitigating the issues in various school administrations. In addition, the approach predicts students with drop-out risk and elements posing a threat to give schools relevant data to customize personalized interventions.

1.3 Research Questions

1. How effective are PAAs in predicting the rate of dropouts?
2. How can PAA predict dropout rates and factors in academic institutions?
3. What is the best predictive algorithm to predict school dropout rates?

1.4 Problem Definition

The elevated drop-out rates are complicated, comprising individual factors and elements unique to the community, various institutions, and learners' groups. Personal and school issues are strongly intertwined; practically all schools in Northern America suffer some student turnover, even the most reputable institutions. Several studies have exposed various risk factors to school drop-out. However, those employing the PAAs approach demonstrate significant variation efficiency in various learning institutions. Therefore, comprehending the efficiency of this approach is critical in ensuring the reduction of drop-out rates.

1.4.1 Problem Scope

While several researchers have extensively documented the causes and impacts of school drop-out, there are few studies on the implementation and effectiveness of PAAs in predicting, monitoring, and addressing the issues. Therefore, this research is to document the efficacy of PAAs in mitigating school drop-outs. The study's scope is limited to North American countries. The approach will be limited to an online survey to review documented materials regarding the issue in the target population.

1.4.2 Dropout Rate Definition

The drop-out rate measures the fraction of students admitted to a specific class or institution in a particular academic term or year who are not actively registered in that grade the following academic year. It is also the proportion of 16- to 26-year-olds that are not recorded in the system and lack a graduate certificate. NCES (2021) defines the term as the rate of leaving a university, college, high school, or any academic institution due to practical reasons such as poor economic status. UNESCO defines drop-out as withdrawing from an education or training program before completion (UNESCO Institute for Statistics, 2022).

1.5 Current Intervention Strategies

Learning institutions have adopted several strategies to contain student drop-outs in various regions. Most focus on improving learner engagement with significant efforts to recognize groups or individuals likely to drop out. Application of educational support during adjustments and analytic monitoring and traceability methods to detect school and student variables that influence drop-out rates are examples of interventions (Aldowah et al., 2020; Faria et al., 2017; Sáiz-Manzanares et al., 2020).

1.5.1 Implement Tracking Systems

Some schools and educational agencies create and manage local information systems to spot future drop-outs, even though there is now only minimal empirical support for longitudinal monitoring systems. Trackers employ unique student identities to enable extensive and longitudinal monitoring of individual students (Faria et al., 2017; Millea et al., 2018). The approach should incorporate a pattern of student absenteeism, poor levels of academic proficiency, grade retention, and detachment from schooling since most scholars link these indicators to higher chances of dropping out.

1.5.2 Provide Academic Enrichment and Support

The approach includes using effective ways to improve academic accomplishment and engage students. These ways incorporate tutoring, assignment aid programs, and intensive academic support delivered as part of the regular academic days, after school, over vacation, or on holidays, as well as other recommended interventions designed to improve student accomplishment (Becker et al., 2017; Hodges et al., 2017). These tactics boost students' motivation, intellectual learning skills, and teaching effectiveness, reducing drop-out likelihood.

1.5.3 Promote Social Skill Development

Some institutions use social skills development programs to address school dropout among students. According to Taylor et al. (2017), the program helps students develop social skills, including problem-solving and effective communication, recognizing, comprehending, moderating emotional reactions, dispute resolution, and goal-setting. It also cultivates appropriate conduct via social skills training to enhance students' sense of identification and connection with the school.

1.5.4 Analytical Tracking Interventions

Some reputable institutions have implemented analytic tracking tools to monitor students with key risk contributions to dropping out and delivering focused educational, social, cultural, and transitory aid for recognized students, which may influence academic achievement and minimize drop-out percentages. The approach emphasizes school-wide activities, establishing realistic and high objectives for all students, and keeping track of trends or progress (Rienties et al., 2017; Xing & Du, 2019). Similar approaches include creating a unified, student-centred environment for learning and creating a supportive culture and caring connections.

1.6 Common Drop-out Risk factors

The central objective of this paper is to examine the efficiency and how PAAs are implemented to mitigate school drop-out rates. However, Carlson (2018) indicates that it is vital to comprehend the risk factors associated with increased drop-out before exploring the preventive mechanisms. As per Ketija (2018), the following are the major risk factors leading to high school dropout rates.

1. Personal backgrounds such as emotional disturbance or learning disability
2. Early adult responsibilities due to poor economic standards in the family
3. Negative social attitudes and peer influence
4. Low social-financial status of the family
5. Students are more socially active outside school than within school
6. Absenteeism, poor class attendance, and low schools engagement
7. Low educational expectations from the student
8. Common education background of parents
9. Lack of extracurricular participation within school.

1.7 Current Dropout Trends

The status drop-out rate (SDR) is the proportion of students between 16 and 24 years old who have not earned any school credential certificate and have not enrolled in any learning institution. There were about 2 million SDRs in 2019, which recorded an SDR of 5.1 percent – shown in figure 1 (NCES, 2021). McFarland et al. (2020) indicate that the SDR in the United States by 2017 was 5.4 percent (as indicated in figure 1 below). The figure also shows that SDR for Bolivian descent, Hispanic subgroup, and Hispanics was 24.4 percent, 1.5 percent, and 9.9 percent, respectively. Figure 2 shows a decline in SDR of about 14.1 percent between 1977 and 2007. There was also a decline of 5.8 percent between 2007 and 2017. In addition, 21.0 percent of SDRs had completed less than nine years of school, 9.8 percent had

completed nine years, 20.3 percent had completed ten years, and 49.0 percent had completed 11 or 12 years in 2017. Between 1977 and 2017 (NCES, 2021), SDR declined 12.0 percent among those who had completed nine years of education and seven percent among those who had completed ten years. Contrary to this, the SDR percentage points rose in those who completed 11 or 12 years by 22.4 % (NCES, 2021). The SDR for students who had completed nine years declined by 7.2 percent. Between 2007 and 2017 (INCES 2021), the performance among those who had completed 11 or 12 years increased by 10 percent. There was no difference among those who had completed ten years.

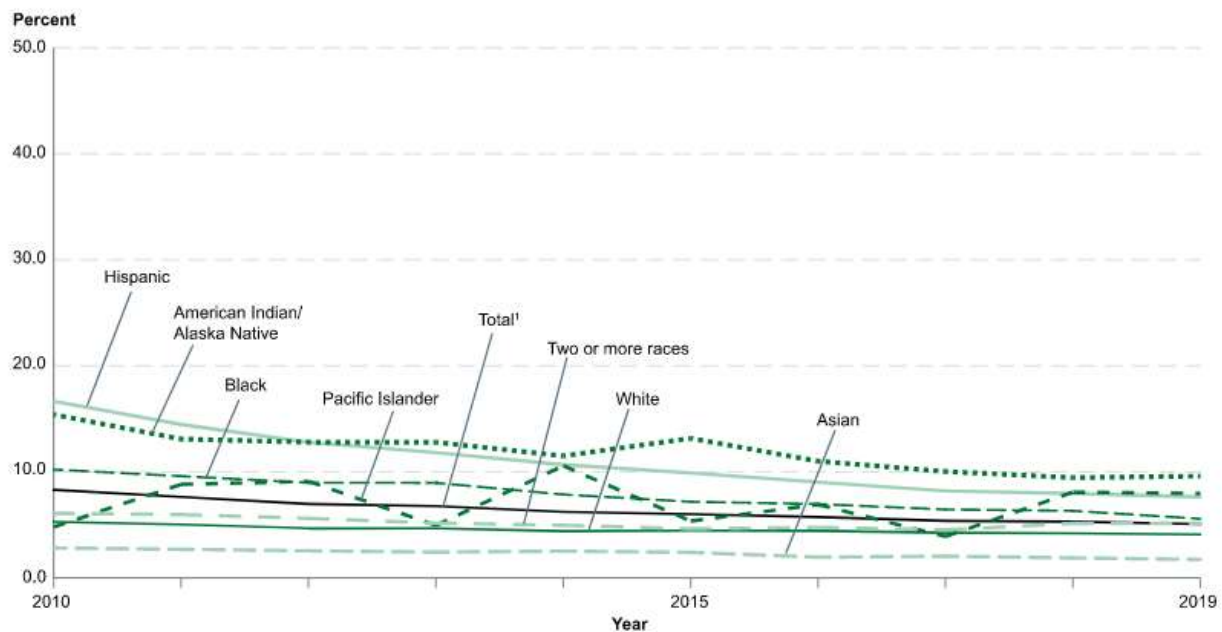


Figure 1: Status Dropout Rate
Source: (NCES, 2021)

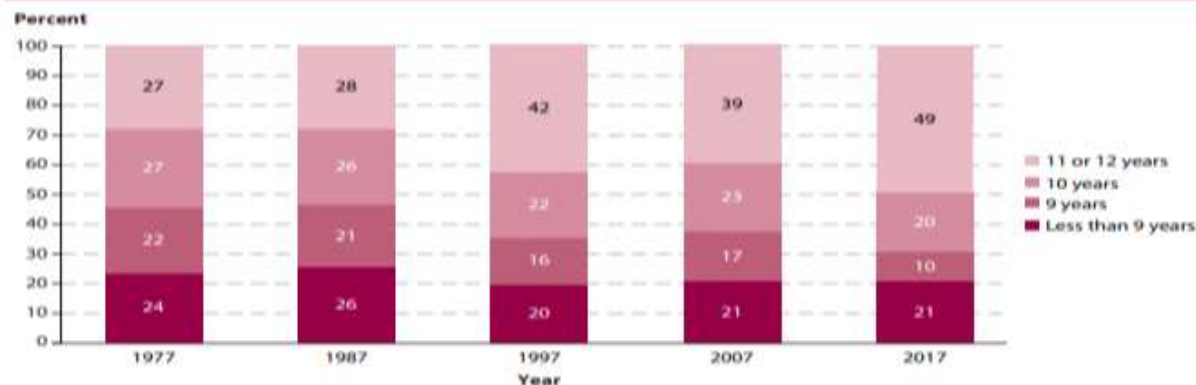


Figure 2: Percentage Distribution of SDR between 1997 and 2017
Source: (McFarland et al., 2018)

1.8 Academic Mediation Theory

The Academic Mediation theory (AMT) elucidates the link between academic success and educational facility dropout. In conventional examinations or grade averages, poor intellectual performance is the most substantial forecaster of students dropping out of school (Ketija, 2018; Balkis, 2018; McFarland et al., 2018). The study suggests that students with low academic success have little school connection and will likely cause dropout from their present educational grade.

1.8.1 General Deviance Theory

The association between aberrant behavior and drop-out rates emphasizes the hypothesis. As per Boyaci (2019), a family's commitment to certain lifestyles and beliefs predicts a child's likelihood of school drop-out. In addition, the theory claims that children's exposure to drug usage during early childhood impacts their development (Ketija, 2018). It also identifies pregnancy as the most common reason for girls dropping out of school.

1.8.2 Deviant Affiliation Theory

This theory emphasizes harmful elements and connections, such as familial abandonment or torture, as causes of dropping out. According to the research, parents, and peers significantly influence students' behavior, growth, and educational accomplishment (Ketija, 2018; Valkov, 2018). At the same time, family bonds would substantially impact how a child socializes with others in the community.

1.8.3 Poor Family Socialization Theory

The low-income family Socialization hypothesis focuses on the family's socioeconomic background. According to Ketija (2018), the family is the first and most essential social structure for individuals and overall growth. Family history significantly impacts a student's attendance and the likelihood of dropping out. Economic inequality is among the primary reasons students feel compelled to leave school to help support their families financially. Other factors include parental separation, parenting stress, family behavioral intention, tolerance level, and the parent's level of education.

1.8.4 Structural Strains Theory

The theory focuses on ethnicity, gender, race, and the family socioeconomic situation surrounding a student. For example, boys and girls are more prone to drop out of school, predominantly if they originate from low-income homes (Ketija, 2018). In addition, boys leave school at a higher rate than girls, owing to their poorer engagement with education and motivation to learn. In contrast, most girls drop out of school due to teenage pregnancy.

1.9 Machine Learning: Contemporary Drop-out Models

School dropout is a severe problem in poor and wealthy countries. Machine learning approaches, on either hand, have gotten much attention as a solution to this problem (Dalipi et al., 2018; Kabathova & Drlik, 2021). On the regular assessment of the dataset for North America as a case study, this research gives a detailed investigation of four supervised classifiers that encompass k-Nearest Neighbor, Neural Networks, Decision Tree, and Naive

Bayes. Furthermore, the study aims to assist researchers in the area with data-driven algorithm suggestions to evaluate how the approach is implemented in the target population.

2. Literature Review

During the 1990s, the issue of student drop-out and retention rose in popularity. Students' age, individual counseling and guidance, campus housing, financial assistance, ethnicity, student gender, student personal traits, and staff gender were all studied in the early research (Wild & Heuling, 2020; Ketija, 2018). Since then, the study has looked at various factors and their impact on retention. Prediction models were created as statistically significant associations were discovered (Agrusti et al., 2019). Besides, to find the best approach to determine students' potential to drop out, researchers utilized a variety of prediction models and compared their findings. Regression analysis is one of the oldest and most widely used approaches for predicting attrition.

Regression analysis is a popular tool for predicting student retention. It has a good track record. It is also one of the mechanisms investigated in this study. In a comparison study using logistic regression, neural networks, decision trees, and their capacity to forecast time to degree accomplishment and retention rates, Hagedorn et al. (2018) used multivariate logistic regression to analyze students' activities during their first and second years. Jaeger and Hinz used logistic regression to predict first-year freshman attrition and student experience with part-time teachers. On the other hand, Yang et al. (2018) used a multinomial model to determine the influence of social integration, motivation, and academic success on student retention.

In recent decades, institutions have used artificial neural networks to categorize and predict outcomes in various domains, such as the analysis of student attrition. Several published publications propose neural network models to forecast retention, course duration, and drop-out. In predicting low, medium, and high drop-out risk or student failure, neural networks outperformed decision trees and discriminant analysis (Imran et al., 2019). Even though none of the outcomes was encouraging, the neural network accuracy rate was equivalent to the other models due to the category rather than the dichotomous prediction strategy. The authors said they might have obtained over 80% accuracy by classifying pupils by chance instead of predicting success or failure.

Mason et al. (2018) shepherded research to study the accuracy of logistic regression, discriminant analysis, neural networks, and structural equation modeling in predicting engineering student retention. They discovered that analysis of variance had the best overall accuracy rate, yet this scored the worst in forecasting vulnerable students, negating its inclusive correctness since the theory's goal is to identify children at risk or forecast non-retained students. Neural networks exhibited the best overall likelihood precision, the probability of recognizing reserved learners, and the possibility of predicting quasi-students among the existing models (Alban & Mauricio, 2019; Mason et al., 2018). Unlike other models, neural networks have demonstrated the highest ability to accept a more extensive data input and infer unseen relationships, thus becoming more accurate (Wu & Feng, 2018). The same approach enhances efficiency in various sectors, including chemical nontoxicity

prediction and earthquake intensity monitoring (Mason et al., 2018). The NN model outperforms all other models, such as K-Nearest Neighbor and decision tree, due to its high accuracy in prediction value, analysis, and classification (Mason et al., 2018). Additionally, it analyses learners' academic progression; identifying learners likely to be promoted to the next class; those likely to continue, and those at risk of dropping out. Educational institutions can use Prediction Alogarithm combined with NN to classify the learners based on their performance to predict the drop-out risk factors.

Baker et al. (2020) try to forecast if individual students would drop out, using data on disciplinary infractions, previous school suspensions, and violence to determine if a student was still on track to finish, along with other consequences. This effort was sluggish in implementation for many years, with just a few initiatives developing each decade (Gubbels et al., 2019). However, many initiatives have surfaced over the last decade that possible effort utilizes predictive analytics and statistical models to predict students quitting school. They use these assumptions as warning metrics (EWIs) to assist public schools in allocating funds to those student groups who are at threat of not graduating (Lee & Chung, 2019; Wood et al., 2017). Data scientists are increasingly working to make these forecasts accessible to many stakeholders, including students, parents, teachers, and community leaders.

Millions of students drop out of school worldwide or are limited from accessing quality education. However, NA is one of the regions with the highest student drop-out rate (McFarland et al., 2020). That is one major factor why the United States was chosen as a case study for this research. In addition, some students are given limited access to elementary and primary education due to various risk factors, such as the family's economic status. Compared to the drop-out issues globally, this study discovered that different nations experience the same problem – the only difference is the problem's depth and scope. For instance, the UK's SDR during 2019 – 2020 was 5.3%, which is 5.1% in the United States (Lough, 2022; NCES, 2021). Each nation has adopted various models to predict the rate of student drop-out and predefined risk factors. Among the models are Decision Tree (DT), Bayesian Networks (BN), Back Propagation Algorithm (BPA), Naive Bayes Model (NBM), Apriori Algorithm model (AAM), Maximum entropy (ME), and Support Vector Machine (SVM) model (Solichin, 2019).

The differential model in *Figure 5* below illustrates the percentage of most common models used worldwide compared with their popularity in NA.

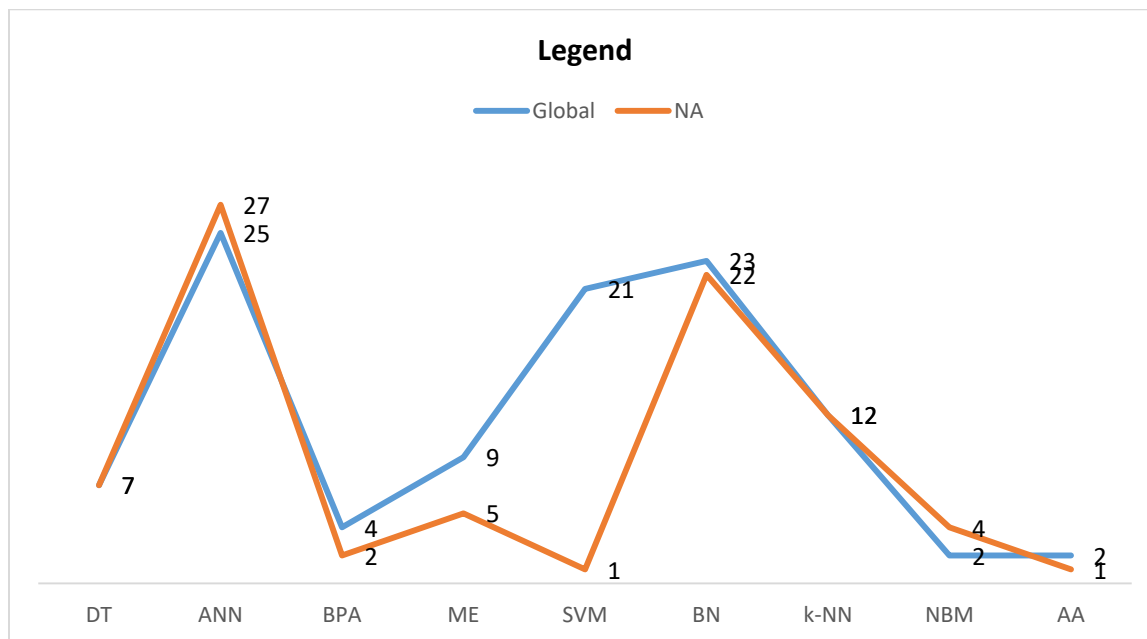


Figure 3: Models used in NA versus globally

Sources: (Nanayakkara et al., 2021; Lough, 2022; NCES, 2021)

From figure 5 above, Bayesian Networks (BN) and Artificial Neural networks (ANN) were the most applied models between 2017 and 2020 for constructing PAAs in NA and worldwide (Lough, 2022). The graph also shows the popularity of adopting k-nearest Neighbors (k-NN) and Decision Tree (DT) (NCES, 2021). Maximum entropy (ME), Decision Trees (DT), and Back Propagation algorithm (BPA) have been used at the intermediate level. In contrast, the Naive Bayes Model (NBM) and Apriori Algorithm have been marginally used (Nanayakkara et al., 2021).

2.1 Confidentiality and Privacy Issues

Models for drop-out prediction are built to accept transformed data into advanced statistical information to enhance inference. The process uses predefined approaches and stores or distributes information to the intended users for analysis. Although collaborative data training breaks data monopoly, it also raises new challenges of model accountability, privacy protection, and data confidentiality (Gu et al., 2019). Although various mechanisms have been implemented to enhance data privacy and confidentiality, data breach incidents expose students' sensitive data (Hlioui et al., 2021). Besides, most government regulations and laws on personal privacy protection do not permit privacy-sensitive and mission-critical domains such as education institutions to share raw data with third parties (Gu et al., 2019). The policy poses a significant challenge in implementing PAAs.

2.2 Relevance of Literature Review

The literature review above was relevant to the study. It provides a solid foundation for comprehending exciting debates and research on the analysis of drop-out rates. It also allows the analysis to build a knowledge base of past initiatives in predicting and mitigating the

issues and the applicability of PAAs in containing future trends. Finally, the review identifies open questions in previous studies and conflicts of interest and addresses research gaps.

2.2.1 Strengths and Weaknesses

The literature review's main strength is that it provides critical knowledge on various mechanisms of managing SDR. However, most of these studies typically focused on investigating the risk factors and possible means of preventing the problem. In addition, no study documents the efficacy of PAA in mitigating school drop-out rates, making it a substantial weakness.

2.3 North American Case Study

The use of predictive analytics strategies (PAAs) within the United States has been sweeping and swift (Bird et al., 2021.). It is estimated that one-third of all institutions have bought into predictive analytics technology and spent hundreds of millions of dollars combined. (Barshay & Aslanian, 2019) While PAAs are cost-effective approaches to managing school drop-outs, are they effective? The following review discusses the findings on whether or not they work to manage school drop-outs effectively and if they are true to life and fair. The author uses North America and the United States interchangeably. In the United States, nearly all educational institutions use predictive analytics (Bird et al., 2021.) The widely used PAAs are most often for identifying which students are least likely to complete their high school education. From this information, positive and negative courses of action take place. The application of PAAs is intended to target individual students who need support, mainly through faculty mentoring. Some institutions in the United States use predictive analytics to highlight courses in which academic achievement is predictive of future favourable outcomes at the school and to earmark students who are anticipated to grapple with specific classes.

In the cases of paid education, however, where tuition is an institution's primary source of revenue, employing predictive analysis models is plausibly used to target students who are most likely to succeed at the school. (Ekowo & Palmer, 2016) In parts of the United States, the practice of PAAs determines which students are predicted to stay in school. For instance, the University of Arizona unearthed from a predictive model that students who earn a C in introductory English writing have a reduced likelihood of graduating. (Treaster, 2017). Schools that want to keep their academic standards high may transfer the student to another institution unofficially. In the circumstances of paid education, some educational institutions will use financial modelling results to direct scholarships to students who are predicted to stay enrolled instead of transferring to another institution. Alternatively, educational institutions will not mentor students to help them attain scholarships, especially those likely to drop out due to a lack of financial support (Ekowo & Palmer, 2016; Paterson, 2019). Not all institutions intervene to help those with academic support needs, a reflection of an earlier unanswered question of whether or not the biases in PAAs are lessened by the broad institutional base from that of online learning.

High drop-out rates have been a severe cultural concern in various learning institutions across North America for many years. Recent studies have suggested using machine learning to develop a drop-out prediction algorithm to monitor, forecast, and mitigate drop-out issues in

learning institutions (Alban & Mauricio, 2019; Xing & Du, 2019). However, no study has explored the effectiveness and efficiency of this approach in achieving its objective. Therefore, this study aims to fill the gap by assessing the approach's efficacy.

2.3.1 Objective

The study's primary objective is to comprehensively review Predictive Analytics Algorithms (PAAs) Strategies and demonstrate how they have been effectively implemented in optimizing retention rates with a case study focusing on the United States.

2.3.2 Target Audience

The target audience for the study is educators, learning institutions, policymakers, government agencies, and the research community.

2.3.3 Problem

School drop-out is a complex problem involving individual factors, the school environment, the family, and the overall community of the student. This issue's complexity makes it more challenging to develop a comprehensive mechanism to contain the problem. Therefore, stakeholders need to implement a sustainable means to manage the issue. One of the possible solutions is utilizing machine learning's PAAs.

2.3.4 Strategies (Classifiers)

The study employed a classifier approach to evaluate the effectiveness of PAAs in mitigating school learning. It utilized four data mining classifiers - k-Nearest Neighbor (k-NN), Neural Networks (NN), Decision Tree (DT), Naive Bayes (NB), and - to categorize student drop rates. Its approach was assessed through a quantitative review to evaluate its effectiveness.

3. Methodology

The study employed a quantitative method to enhance the objective measurement of the available data's numerical, mathematical, and statistical analysis. This approach aims to receive a more precise measurement of the efficacy of the PAA and how it has been effectively implemented through preexisting literature. As a result, the study employed a thematic study design. It reviewed over 500 articles relating to the study topic. This figure was reduced to 40 articles after conducting a thematic content assessment.

3.1 Data Analysis

The paper utilized DT, k-NN, NB, and NN data mining classifiers as reference points during data analysis. The classifiers applied nine student variables as the primary input. The variables were; parent and student's educational level, age, gender, previous learning experience, school engagement, social environment, extracurricular participation, peer influence, and family's socio-economic status. In addition, the study organized data to remove extreme outliers and inaccurate records. Finally, it utilized excel and SPSS for data computation and representation.

4. Results

The study adopted data mining classifiers to classify student drop-outs to develop a hypothesis on the efficacy of PAAs. Berrar (2019) 's 10-fold approach is employed to validate each classifier and measure the average performance to measure the accuracy of each model. Receiver operating characteristics (ROC) analysis was also used to evaluate the predictive models to compare the accuracy between the models and test data. A return of true and false rates demonstrates the independence and performance of the model (Lee & Chung, 2019). Finally, data obtained from ROC analysis were used to predict the significance and efficiency of the model in predicting student dropouts. As indicated in Table 1 below, three variables, social environment, peer influence, and parent's socio-economic status, were found in most PAAs to be the most predominant causes of school drop-out.

Table 1: Predominant Risk Factors Based on PAAs

PAAs Classifiers	NB	NN	DT	3-NN
Variables	Social environment	Peer influence	Peer influence	Peer influence
	Socioeconomic	Social	Social	Age
	status of the family	environment	environment	
	School engagement	Socioeconomic	Socioeconomic	Gender
		status of the	status of the	
		family	family	
	Extracurricular	Previous	Extracurricular	Educational
	participation	learning	participation	level
		experience		

Source: (Lee & Chung, 2019)

The review of available literature designates the varying efficiency of data mining classifiers in terms of accuracy, classification, and data sensitivity (See Figure 4). NB classifier demonstrated better precision and classification results than the other two (Ba'abbad et al., 2021; Ababneh, 2019). It recorded approximately 4% higher than 3-NN, DT, and NN. The remaining showed varied efficiency (Ba'abbad et al., 2021). However, they all recorded a minimum accuracy rate of 81%, which indicates that PAAs can be the best approach to predicting school drop-outs (Ababneh, 2019).

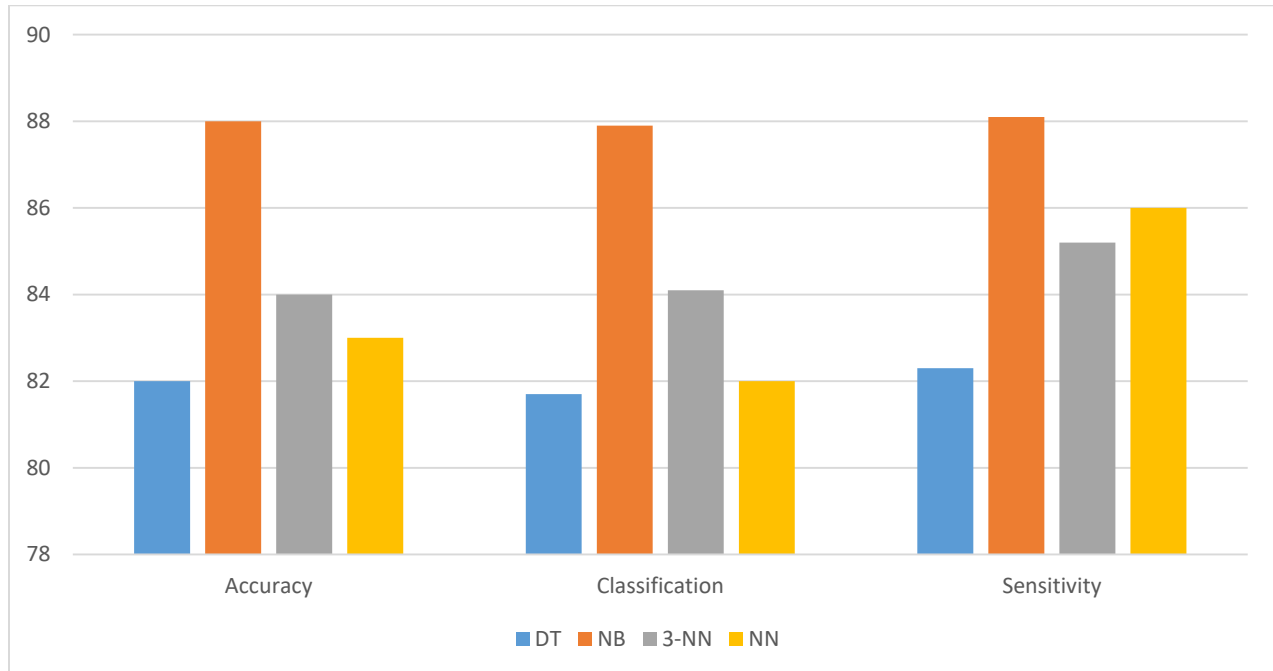


Figure 4: Efficiency of classifiers

Source: (Ba'abbad et al., 2021; Ababneh, 2019)

5. Discussion

This study aimed to explore how effective PAAs can be implemented in the education sector to optimize student retention rates in North America. The study themes revolved around exploring common drop-out risk factors, current trends in school drop-out in NA, the best PAAs for the problem, and how effective PAAs are in addressing the issues. It achieved its objectives by reviewing available literature regarding its themes. Besides, it discovered that four classifiers – DT, k-NN, NB, and NN – are predominantly used in most institutions for data predictions. These models usually employ drop-out risk factors, including parent and student's educational level, age, gender, previous learning experience, school engagement, social environment, extracurricular participation, peer influence, and family's socio-economic status.

Although numerous studies have demonstrated the feasibility of using PAAs in predicting and mitigating student drop-out rates, very few have tried to measure the approach's effectiveness. Therefore, applying data mining techniques to measure the validity of predicting algorithms was a substantial venture for the study. The study measured the feasibility of student attrition and clarified the parameters advantageous in the prediction using the proper model. It designates that a computational intelligence strategy combines many classifiers to enhance prediction accuracy. These classifiers and models include DT, k-NN, NB, and 3-NN. However, NB demonstrated a higher accuracy rate than other classifiers in solving the issue in complex scenarios. In addition, it can learn and model complex and nonlinear relationships.

Additionally, the study learned that most institutions design PAAs models to receive preprocessed data and transform them into advanced statistical information to improve inference. They collect user data, perform analyzes using preset methods, and then store or deliver it to the relevant users for analysis. However, the study has realized a new challenge from this approach. The procedure introduces new challenges, such as anonymity, privacy, and confidentiality of student data. Although numerous measures have been adopted to improve data privacy and confidentiality, data breaches still occur, exposing students' sensitive data. The situation is a significant obstacle to PAA implementation. The same approach (data mining) can be adapted to detect and prevent student data from leaking.

The results provide an appropriate and compelling response to the research questions. The breakdown of this response includes identifying common drop-out risk factors in the US, including the effectiveness of PAAs in predicting drop-out rates; common drop-out trends were also identified in addition to identifying the most appropriate model.

The implications of drop-out models in maximizing academic institution retention rates cannot be adequately emphasized. These models, when effectively applied, can contribute significantly to addressing one of the major deterrents in maintaining student continuity through graduation in any institution. Hence the future benefits of using predictive models in these institutions outweigh the risks implicitly and explicitly. Moreover, given the current trend, there is every reason that model applications will continue to motivate academic institution managers, administrators, and policymakers attempts to improve attrition and compelling management concerns. It is also appropriate to highlight how institutions have used these outcomes to improve student dynamics. Most institutions have used models to improve attrition rates, stabilize revenues, and adapt course developments.

An invaluable theme and limitation of the article are limited or inadequate access to details of relevant data management (DM) procedures and transformations. DM is an essential component of any data-driven system and must be prioritized, highlighted, and strengthened. One missing thematic strategy that requires more consideration and inclusion in implementations like PAAs is the introduction of mixed-method techniques.

While the current quantitative application of classifiers serves and provides beneficial outcomes, the complementary inclusion of qualitative methods contributes substantially to enhancing the results, producing more comprehensive, compelling, and meaningful conclusions and outcomes. For example, assessing the impact and contribution of crucial risk factors, like bullying, peer pressure, and others, must be part of the analytic process. Omitting such variables is a missing link that is not critical but remains plausible and vital in addressing academic institution drop-out dynamics.

5.1 Study Challenges

The main challenge which affected the study was the incompleteness of existing data. In addition, the research relied on a secondary data collection approach. As a result, it was not easy to obtain specific data from one depository. Besides, it was challenging to establish data accuracy, sensitivity, and generality of data across North America learning institutions.

5.2 Research Contribution to Current Initiatives

While there are numerous studies on various approaches to using data mining and machine learning to develop PAAs to predict risk factors associated with student drop-out, there is limited research on how the process is implemented and its efficiency. Therefore, this study will significantly contribute to the existing literature in various ways. First, it will close the gaps in the efficiency of PAAs models and classifiers on the issues. It will also create a foundation for understanding the best predictive algorithm learning institutions can adopt to achieve accuracy. Finally, this study will allow scholars to comprehend how PAAs can optimize school retention rates effectively.

5.2.1 How Research Question has been Answered

The research focused on identifying common trends in drop-out in North American schools, the common drop-out risk factors, and the effectiveness of PAAs in predicting the attrition rate. It was also to determine how PAA predicts the drop-out rates and factors influencing it and to identify the best predictive algorithm to predict school drop-out rates. The researcher identified common trends such as student drop-out and retention rates. Common drop-out risk factors include age, gender, past learning experiences, school engagement, social environment, peer influence, parent and child educational level, and family socio-economic factors (Lee & Chung, 2019). PAA identified various data mining models and classifiers such as Neural Networks (NN), k-Nearest Neighbour (k NN), Decision Tree(DT), and Naïve Bayes (NB) (Lee & Chung, 2019). They were used to illustrate how effective they are in predicting analytics in most academic institutions in North America. In a study to determine the most effective approach among the four models in analyzing the common drop-out risks, NB was considered the most effective due to its ability to analyze complex data (Ba'abbad et al., 2021; Ababneh, 2019). It was done using the Receiver Operating Characteristics (ROC) to evaluate the effectiveness of the models and the test data (Lee & Chung, 2019). Therefore, many institutions in NA have embraced the PAA strategy to retrieve, analyze and store their statistical data.

5.3 Implications and Future dynamics of PAA Applications on Academic Institution Management

As indicated above, PAA has significant future implications, especially in managing academic institutions. Predictive analytics will enable institutions to identify trends and changes in the educational structure (Blazer &Gonzalez, 2018). They will analyze graduation rates, program enrollments, library records, learners' performance, and attendance. PAA is an effective tool for identifying students at drop-out risk using models such as Neural Networks to analyze and classify the positive and negative predictive values (Alaban &Mauricio, 2019). Through this, managers in academic institutions can classify the well-performing, average, and poorly-performing students to help them determine the strategies to employ to help the low-performing students reduce the school drop-out rate (Blazer &Gonzalez, 2018). Analyzing the academic institutions' data will also optimize student retention since their instructors will guide them on objective approaches to ensure they perform well in their grades. Additionally, PAA can be helpful in academic institutions since it can enable

managers and curriculum developers to adjust the curriculum based on compelling and relevant factors without compromising the quality of the course.

5.4 How Research Findings Have Benefitted Academic Institutions

Based on the research, most academic institutions have benefitted from the PAA approach. Study shows that 33 % of institutions in the united stated have incorporated this machine learning to analyze and classify their data to mitigate the increased school drop-out rates (Barshay & Aslanian, 2019). Through the help of specific faculties, academic institutions collect data using predictive analysis to classify students with good grades and those with poor ones who need support (Balkis, 2018). In addition, they use predictive analysis to modify their course to guide new students on handling coursework to maximize the retention rate. Moreover, in those academic institutions paying tuition fees, the financial managers use PAA to determine whether students remain in school and access learning and will not access knowledge.

6. Conclusion

This paper aims to conduct a compressive appraisal of PAAs strategies and demonstrate how they have been effectively implemented to mitigate institutional drop-out in North America (Blazer & Gonzalez, 2018). The study established that predicting and addressing retention rates is challenging for educators, the government, policymakers, and the community. Therefore, it utilized a secondary data collection method through a comprehensive literature review. The efficacy was measured using data mining classifiers, including DT, k-NN, NB, and NN, to measure the effectiveness and implementation of the approach. It was also observed that the method had been utilized by various educational institutions across North America, using a combination of multiple predicting algorithms to ensure efficiency. In addition, the study shows that each student drop risk factor is vital in predicting the model's outcome.

Finally, "no child left behind" is a constant reminder to all policymakers nationally, regionally, and globally (Rienties et al., 2017). The challenges of addressing drop-out rates are multi-dimensional, complex, and inclusive. No nation is exempt. While the advances in information technology – PAAs - have significantly contributed to intervention strategies, there is evidence that approaches come with their strengths and weaknesses. Administrators must consider these challenges seriously and develop more dynamic solutions, especially refining more compelling processes. Effective and sustainable solutions should be processes and not events. The benefits of successful outcomes are significant and cannot be adequately emphasized. They range from improved quality of life for the population to strengthened economic power of a given nation.

References

Agrusti, F., Bonavolontà, G., & Mezzini, M. (2019). University dropout prediction through educational data mining techniques: A systematic review. *Journal of e-learning and knowledge society*, 15(3), 161-182. <https://doi.org/10.20368/1971-8829/1135017>

- Alban, M., & Mauricio, D. (2019). Neural networks to predict dropout at universities. *International Journal of Machine Learning and Computing*, 9(2), 149-153.
https://web.archive.org/web/20200307075313id_/http://www.ijmlc.org/vol9/779-ML0074.pdf
- Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., & Alalwan, N. (2020). Factors affecting student dropout in MOOCs: a cause and effect decision-making model. *Journal of Computing in Higher Education*, 32(2), 429-454. <https://doi.org/10.1007/s12528-019-09241-y>
- Ba'abbad, I., Althubiti, T., Alharbi, A., Alfarsi, K., & Rasheed, S. (2021). A Short Review of Classification Algorithms Accuracy for Data Prediction in Data Mining Applications. *Journal of Data Analysis and Information Processing*, 9(3), 162-174. <https://doi.org/10.4236/jdaip.2021.93011>
- Baker, R. S., Berning, A. W., Gowda, S. M., Zhang, S., & Hawn, A. (2020). Predicting K-12 dropout. *Journal of Education for Students Placed at Risk (JESPER)*, 25(1), 28-54. <https://doi.org/10.1080/10824669.2019.1670065>
- Balkis, M. (2018). Academic motivation and intention to school dropout: the mediation role of academic achievement and absenteeism. *Asia Pacific Journal of Education*, 38(2), 257-270. <https://doi.org/10.1080/02188791.2018.1460258>
- Barshay, J., & Aslanian, S. (2019). Under a watchful eye: Colleges use big data to track students to boost graduation rates, but it comes at a cost (APM Reports). <https://www.apmreports.org/story/2019/08/06/college-data-tracking-students-graduation>
- Becker, M. A. S., Schelbe, L., Romano, K., & Spinelli, C. (2017). Promoting first-generation college students' mental well-being: Student perceptions of an academic enrichment program. *Journal of College Student Development*, 58(8), 1166-1183. <https://doi.org/10.1353/csd.2017.0092>
- Berrar, D. (2019). Cross-Validation. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Bird, Castleman and Mable (2021) Bringing Transparency to Predictive Analytics: A Systematic Comparison <http://journals.sagepub.com> January-December 2021, Vol. 7, No. 1, pp. 1-19 DOI: 10.1177/23328584211037630
- Blazer, C., & Gonzalez Hernandez, V. (2018). Student Dropout: Risk Factors, Impact of Prevention Programs, and Effective Strategies. Research Brief. Volume 1708. Research Services, Miami-Dade County Public Schools. <http://files.eric.ed.gov/fulltext/ED587683.pdf>
- Boyaci, A. (2019). Exploring the Factors Associated with the School Dropout. *International Electronic Journal of Elementary Education*, 12(2), 145-156. <https://www.iejee.com/index.php/IEJEE/index>

- Carlson, S. E. (2018). Identifying students at risk of dropping out: indicators and thresholds using ROC analysis. <https://digitalcommons.georgefox.edu/edd/114/>
- Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, 135, 113325. <https://doi.org/10.1016/j.dss.2020.113325>
- Dalipi, F., Imran, A. S., & Kastrati, Z. (2018, April). MOOC dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE global engineering education conference (EDUCON)* (pp. 1007-1014). IEEE. <https://doi.org/10.1109/EDUCON.2018.8363340>
- Ekowo, M., & Palmer, I. (2016, October). The promise and peril of predictive analytics in higher education: A landscape analysis (New America Policy Paper). <https://www.newamerica.org/education-policy/policy-papers/promise-and-peril-predictive-analytics-higher-education/>
- Evans, W. N., Kearney, M. S., Perry, B., & Sullivan, J. X. (2020). Increasing Community College Completion Rates Among Low-Income Students: Evidence from a Randomized Controlled Trial Evaluation of a Case-Management Intervention. *Journal of Policy Analysis and Management*, 39(4), 930-965. <https://doi.org/10.1002/pam.22256>
- Faria, A. M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). Getting Students on Track for Graduation: Impacts of the Early Warning Intervention and Monitoring System after One Year. REL 2017-272. *Regional Educational Laboratory Midwest*. <http://files.eric.ed.gov/fulltext/ED573814.pdf>
- Gkontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2019). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 1-16. <https://doi.org/10.1080/10494820.2019.1709209>
- Gu, Z., Jamjoom, H., Su, D., Huang, H., Zhang, J., Ma, T., ... & Molloy, I. (2019, June). Reaching data confidentiality and model accountability on the Caltrain. In *2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* (pp. 336-348). IEEE. <https://doi.org/10.1109/DSN.2019.00044>
- Gubbels, J., van der Put, C. E., & Assink, M. (2019). Risk factors for school absenteeism and dropout: a meta-analytic review. *Journal of youth and adolescence*, 48(9), 1637-1667. <https://doi.org/10.1007/s10964-019-01072-5>
- Hagedorn, L. S., Maxwell, W., & Hampton, P. R. E. S. T. O. N. (2019). Correlates of retention for African-American males in community colleges. In *Minority Student Retention* (pp. 7-27). Routledge. <http://interwork.sdsu.edu/sp/m2c3/wp-content/blogs.dir/2/files/2012/10/Hagedorn-Maxwell-Hampton-2001-02.pdf>

- Hlioui, F., Aloui, N., & Gargouri, F. (2021). A Withdrawal Prediction Model of At-Risk Learners Based on Behavioural Indicators. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 16(2), 32-53. <https://www.igi-global.com/journal/international-journal-web-based-learning/1081>
- Hodges, J., McIntosh, J., & Gentry, M. (2017). The effect of an out-of-school enrichment program on the academic achievement of high-potential students from low-income families. *Journal of Advanced Academics*, 28(3), 204-224. <https://doi.org/10.1177%2F1932202X17715304>
- Imran, A. S., Dalipi, F., & Kastrati, Z. (2019, April). Predicting student dropout in a MOOC: An evaluation of a deep neural network model. In *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence* (pp. 190-195). <https://doi.org/10.1145/3330482.3330514>
- Kabathova, J., & Drlik, M. (2021). Towards predicting student dropout in university courses using different machine learning techniques. *Applied Sciences*, 11(7), 3130. <https://doi.org/10.3390/app11073130>
- Kemper, L., Vorhoff, G., & Wigger, B. U. (2020). Predicting student dropout: A machine learning approach. *European Journal of Higher Education*, 10(1), 28-47. <https://doi.org/10.1080/21568235.2020.1718520>
- Ketija, D. (2018). Reasons behind children dropping out of primary schools with unequal socio-economic preconditions: A comparative case study of three primary schools in Babati Town Council, Tanzania. <https://www.diva-portal.org/smash/get/diva2:1215119/FULLTEXT01.pdf>
- Ketija, D. (2018). Reasons behind children dropping out of primary schools with unequal socioeconomic preconditions: A comparative case study of three primary schools in Babati Town Council, Tanzania. <https://www.diva-portal.org/smash/get/diva2:1215119/FULLTEXT01.pdf>
- Lee, S., & Chung, J. Y. (2019). The machine learning-based dropout early warning system improves the performance of dropout prediction. *Applied Sciences*, 9(15), 3093. <https://doi.org/10.3390/app9153093>
- Lough, C. (2022, March 17). Degree course dropout rates fall to the lowest on record. Independent. <https://www.independent.co.uk/news/uk/university-of-cambridge-university-of-oxford-russell-group-higher-education-statistics-agency-open-university-b2037939.html>
- Marôco, J., Assunção, H., Harju-Luukkainen, H., Lin, S. W., Sit, P. S., Cheung, K. C., ... & Campos, J. A. (2020). Predictors of academic efficacy and dropout intention in university students: Can engagement suppress burnout?. *PloS One*, 15(10), e0239816. <https://doi.org/10.1371/journal.pone.0239816>
- Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a

- backpropagation neural network and logistic regression. *Research in Higher Education*, 59(3), 382-400. <https://doi.org/10.1007/s11162-017-9473-z>
- McFarland, J., & Cui, J., Rathbun, A., & Holmes, J. (2020). Trends in High School Dropout and Completion Rates in the United States: 2018. Compendium Report. NCES 2019-117. National Center for Education Statistics. <https://nces.ed.gov/pubs2020/2020117.pdf>
- McFarland, J., Cui, J., Rathbun, A., & Holmes, J. (2018). Trends in High School Dropout and Completion Rates in the United States: 2018. Compendium Report. NCES 2019-117. *National Center for Education Statistics*. <https://nces.ed.gov/pubs2020/2020117.pdf>
- Millea, M., Wills, R., Elder, A., & Molina, D. (2018). What matters in college student success? Determinants of college retention and graduation rates. *Education*, 138(4), 309-322. <https://www.ingentaconnect.com/content/prin/ed/2018/00000138/00000004/art00003>
- Mohamed, M. H., & Waguih, H. M. (2018). A proposed academic advisor model based on data mining classification techniques. *International Journal of Advanced Computer Research*, 8(36), 129-136. <http://dx.doi.org/10.19101/IJACR.2018.836003>
- Nanayakkara, A. C., Kumara, B. T. G. S., & Rathnayaka, R. M. K. T. (2021). A Survey of Finding Trends in Data Mining Techniques for Social Media Analysis. *Sri Lanka Journal of Social Sciences and Humanities*, 1(2). <http://doi.org/10.4038/sljssh.v1i2.36>
- NCES. (2021). Dropout Rates. National Center for Education Statistics. Fact Sheets. <https://nces.ed.gov/fastfacts/display.asp?id=16>
- Rastrullo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance using machine learning: A review. *Applied sciences*, 10(3), 1042. <https://doi.org/10.3390/app10031042>
- Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a learning analytics intervention and evaluation framework: What works? *Big data and learning analytics in higher education* (pp. 147-166). Springer, Cham. https://doi.org/10.1007/978-3-319-06520-5_10
- Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., & García-Osorio, C. I. (2020). Monitoring students at the university: Design and application of a moodle plugin. *Applied Sciences*, 10(10), 3469. <https://doi.org/10.3390/app10103469>
- Solichin, A. (2019, September). Comparison of Decision Tree, Naïve Bayes, and K-Nearest Neighbors for Predicting Thesis Graduation. In *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 217-222). IEEE. <http://doi.org/10.23919/EECSI48112.2019.8977081>

- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88(4), 1156-1171. <https://doi.org/10.1111/cdev.12864>
- Treaster, J. B. (2017, February 2). Will you graduate? Ask big data. The New York Times. https://www.nytimes.com/2017/02/02/education/edlife/will-you-graduate-ask-big-data.HTML?_r=1
- UNESCO Institute for Statistics. (2022). UIS Glossary: Education. UNESCO UIS. https://uis.unesco.org/sites/default/files/documents/uis_glossary_education_20200921.pdf
- Valkov, P. (2018). School dropout and substance use: Consequence or predictor. *Trakia Journal of Sciences*, 16(2), 95. <https://pdfs.semanticscholar.org/7bee/306c6f4b286993be90f35151a80af1d12a9c.pdf>
- Wild, S., & Heuling, L. S. (2020). Student dropout and retention: An event history analysis among students in cooperative higher education. *International Journal of Educational Research*, 104, 101687. <https://doi.org/10.1016/j.ijer.2020.101687>
- Wood, L., Kiperman, S., Esch, R. C., Leroux, A. J., & Truscott, S. D. (2017). Predicting dropout using student-and school-level factors: An ecological perspective. *School Psychology Quarterly*, 32(1), 35. <https://psycnet.apa.org/record/2016-15743-001>
- Wu, Y. C., & Feng, J. W. (2018). Development and application of artificial neural networks. *Wireless Personal Communications*, 102(2), 1645-1656. <https://doi.org/10.1007/s11277-017-5224-x>
- Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3), 547-570. <https://doi.org/10.1177/2F0735633118757015>
- Yang, S. J., Lu, O. H., Huang, A. Y., Huang, J. C., Ogata, H., & Lin, A. J. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. *Journal of Information Processing*, 26, 170-176. <https://doi.org/10.2197/ipsjip.26.170>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)

