World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Review Article)

Comprehensive analysis of factors influencing the real-world application of machine learning for student success rate calculation and their impacts on student achievement & educational institutions

Divvyam Arora *

High School Senior STEM (Scholar Badge) & President, Physics Club (Redshift), Delhi Public School (DPS), Sector 45, Gurgaon, Haryana, India.

World Journal of Advanced Research and Reviews, 2023, 19(03), 942–953

Publication history: Received on 30 June 2023; revised on 17 September 2023; accepted on 19 September 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.19.3.1879

Abstract

The successful application of machine learning for student success rate calculation is contingent upon a comprehensive analysis of several interrelated factors that span the educational landscape. Educational institutions must ensure that they have access to accurate and relevant data, including student demographics, academic performance records, and even extracurricular activities, to build effective predictive models. Moreover, dataprivacy and security concerns must be addressed to maintain the integrity of sensitive student information. Next, the choice of machine learning algorithms is crucial. The selection should align with the specific objectives of predicting student success, taking into account the type and volume of data available. Factors such as classification algorithms, regression techniques, and deep learning models must be carefully considered, and their performance must be assessed through rigorous testing and validation. Addressing potential biases in machine learning models is crucial to ensure equitable outcomes. Careful attention must be paid to the training data, as biased data can lead to discriminatory predictions. Ongoing monitoring and model refinement are necessary to minimize these biases and promote fairness in student success predictions. Thus present paper is focused on a comprehensive analysis of factors influencing the real-world application of machine learning for student success rate. In order to achieve this goal, relevant researches on machine learning is considered and further enhancement has been made to make the proposed work more efficient.

Keywords: Machine learning; Student success rate; Education; Data privacy

1. Introduction

Several interconnected aspects throughout the educational environment must be analyzed in depth for machine learning to be successfully applied to the computation of student success rates. In order to construct reliable prediction models, educationalinstitutions must have access to complete and accurate data on student characteristics, academic achievement, and extracurricular activities. Additionally, data privacy and security issues must beresolved to protect students' personal information. The next most important step is picking the right machine learning algorithms [1]. The goals of forecasting student achievement should inform thechoices, as should the kind and amount of data at hand. Classification algorithms, regression methods, and deep learning models are just a few examples of important factors that need to be thought out andtested thoroughly to determine how effective theywill be. It is critical to address any biases in machine learning algorithms to provide fair results. Training data should be scrutinized sincediscriminatory predictions might result from usinginaccurate or incomplete data [2]. Fairness in predicting students' future successes requires constantmonitoring and model updating to reduce inherent biases.

* Corresponding author: Divvyam Arora.

1.1. Machine learning

Machine learning (ML) is an umbrella term forsolving problems for which the development of algorithms by human programmers would be prohibitively expensive; instead, these problems are solved by guiding machines to "discover" their "own" algorithms, without being given any explicit instructions from any human-developed algorithms. Generative artificial neural networks have recently been able to outperform several older methods.

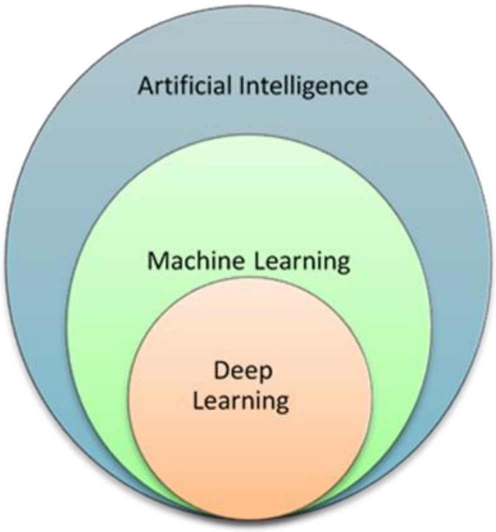

Figure 1 Machine Learning

In fields where it would be too expensive to design algorithms to handle the necessary tasks, such as huge language models, computer vision, voice recognition, email filtering, agriculture, and health, machine-learning techniques have been utilized. Mathematical optimization (mathematical programming) techniques are the theoretical underpinnings of ML. Data mining is a similar (parallel) topic of research that uses unsupervised learning to do exploratory data analysis [2-4].

1.2. Student Success Rate Calculation

The definition of "success" and the uniquecircumstances in which the success rate is being calculated may have a significant impact on the results. There are several methods to assess academicachievement, so defining your criteria in advance of arate-of-success analysis is essential. Here are some typical indicators and procedures for determining student completion rates:

1.2.1. Define Success Criteria

Determine what you consider as a successful outcome for students. Common success criteria include:

- Graduation or completion of a specific program or course.
- Achieving a certain grade pointaverage (GPA).
- Passing specific exams or assessments.
- Gaining employment in a relevantfield after graduation.

1.2.2. Gather Data

Collect the necessary data to assess student outcomes. This data may come from various sources, such as student records, transcripts, exam results, and employment records.

1.2.3. Calculate Success Rates

Depending onyour defined criteria, you can calculate success rates using the following formulas:

1.2.4. Graduation Rate

Graduation Rate = (Number of Graduates / Total Number of Students) * 100

1.2.5. Course Completion Rate

Course Completion Rate = (Number of Students Who Completed the Course / Total Enrolled in theCourse) * 100

1.2.6. GPA-Based Success Rate

Determine a GPA threshold for success (e.g., 3.0 or higher) and calculate the percentage of studentswho meet or exceed this threshold.

1.2.7. Exam Pass Rate

Exam Pass Rate = (Number of Students Who Passedthe Exam / Total Number of Students Who Took the Exam) * 100

1.2.8. Employment Success Rate

Employment Success Rate = (Number of Graduates Employed in relevant Field / Total Number of Graduates) * 100

1.2.9. Analyze and Interpret the Data

Once youhave calculated the success rates, analyzethe results to gain insights into the performance of your students. Look for trends, patterns, and areas for improvement.

1.2.10. Consider Context

Keep in mind that success rates can vary depending on factors like the type of institution, program, student demographics, and external factors. It's essential to interpret the results in thecontext of these variables.

1.2.11. Continuous Improvement

Use the data to identify areas where interventions may be needed to improve student success rates. Implement strategies to address any challenges or issues that emerge from the analysis.

1.2.12. Monitor Progress

Regularly track and update success rates to assess the impact of any interventions or changes implemented toimprove student outcomes.

1.3. Machine Learning in Data Privacy

When it comes to protecting personal informationwhile yet allowing for important insights to be derived from datasets, machine learning plays a crucial role. Differential privacy techniques are a crucial use of machine learning in the realm of data privacy. These methods introduce random variations into the data before analysis, protecting the privacy of individuals while maintaining the integrity of the dataset as a whole. Moreover, machine learning models may be used for anomaly detection, which might reveal data breaches or unauthorized access. In addition, organizations may work together and exchange insights without disclosing sensitive data thanks to privacy-preserving machine learning techniques that allow the building of models on encrypted or privacy-enhanced data. In conclusion, machine learning not only protects personal information but also allows businesses to use data to make more informed decisions in compliance with legal mandates [5].

1.3.1. Role of Machine Learning Used StudentSuccess Rates in Education [6, 7]

When it comes to assessing and enhancing student success rates in education, machine learning plays a significant role while also addressing the critical topic of data privacy. Academic achievement records,attendance records, engagement data, and socioeconomic characteristics may all be analyzed with the use of machine learning algorithms in today's schools. These algorithms may uncoverpatterns and trends that teachers would miss, shedding light on the underlying causes of students' successes and failures. Additionally, machine learning algorithms can forecast students' performance, allowing schools to take preventative measures. For instance, they may help teachers pick out children who are at danger of falling behind or dropping out of school, so that they can give timely assistance and individualized interventions to boost those students' chances of succeeding. Individual students may profit from this predictive capacity, but institutions can also use it to better allocate resources and fine-tune their curricula. However, it is essential to strike a balance between these developments and concerns about personal data privacy. Information about students is private and must be protected. Encryption, access limits, and anonymization methods are just some of the security measures that schools should put in place to secure students' personal information. The Family Educational Rights and Privacy Act (FERPA) in the United States regulate the use and disclosure of student data, and shall be followed [8-12].

2. Literature Review

K. Al Mayahi et al. (2020) focused on the Student's Performance in School Using Machine Learning. Today, all organizations were hastened their use of AI tools in order to see clearly and make progress. The availability of such large datasets has allowed forthe use of AI in many fields, included education. In this study, they developed an automated learning model to forecast a learner's academic success. The proposed model used historical student data and final exam results. Adoption of the model was backed by ahigh rate of accuracy. [1].

M. Anusha et al. (2019) presented machine learning for predicted students' grades. In this work, they examine the literature on predicted student performance with the aim of identified underachiever.Based on the result of the study, several machine learning approaches were used to address issues with risk prediction and evaluation of student performance. The student performance prediction system may be greatly enhanced by make use of machine learning method to track and forecast student growth and achievement. [2]

C. J. Arizmendi et al. (2022) reviewed computerized records of student behavior in the classroom for predictive purposes. They introduce feature mapped and analysis of LMS data as a means of achieving this aim. Then begin with a literature review that provides a concise summary of the various methods currently began used to analyze LMS data, and they proceed to examine privacy concerns raised by the use of demographic data, as well as questions of fair model construction. The paper's second section provides an introduction of well-known ML algorithms and discusses analytical factors such feature creation, model performance evaluation, and sampled strategies. In the end, they provide an empirical example to show how accurate predictions may be made using LMS data, highlighting key aspects and evaluating model performance across a range of model setting. [3]

R. B. Saglam et al. (2022) explained people see, categorize, and develop knowledge about themselves.Therefore, the goal of this work was to considerably enhance research in this area by analyzing the treatment of personal data in many contexts, including laws and regulations, application privacy policies, and scholarly papers. This study adds to the understand of the basic disconnects existing and also the social consequences of new technologies by an evaluation of how personal information has developed and was interpreted differently among these important stakeholders. To further aid academics and practitioners in their efforts to work with or build solutions to preserve such information, they present a set of unique taxonomies of personal information [4].

T. Cardona et al. (2023) reviewed the model for learning retention in higher education using data mining and machine learning. The important results, such as the identification of the components used in previous studies and approaches employed for prediction, present a study viewpoint on the topic of student retention prediction using machine learning. These results might be utilized to design even more extensive research, increase the accuracy of predictions, and, by extension, inform strategies to boost student retention [5].

K. Hamann et al. (2021) focused on the issues of retention and achievement in online political science courses. They investigate how students' choice of course delivery methods influences their academicoutcomes. The data show that students generally do worse as they increase their reliance on online courses. When responding to this shift in higher education, institutions must pay close attention to the needs of distance learner. The COVID-19 epidemic has driven a shift to online education throughout the United States and worldwide, and this trend is expected to continue into the foreseeable future, making it all the more important to take into account the demands of online learners [6].

B. Holicza et al. (2023) did research on machine learning algorithms to predict and compare students' online and offline academic performance. In this study, they use several machine learning methods, such as the support vector machine with different kernels, the decision tree, the random forest,and the k-nearest neighbor's algorithm, to make predictions about student performance and to evaluate these predictions. Predicted vulnerabilitiesare compared with metrics like F1 score and accuracy, and they also compare two databases, one with data linked to online learning and another with data on important offline learning aspects. However, the databases need to be normalized to suit the forecast format before the algorithms can be used. They conclude that behaviors like sleep, study, and use of electronic media all have a role in academicachievement. This publication elaborates on thefindings in further depth [7].

S. Kaddoura et al. (2022) presented machine learning methods for online education and testing. After looking at 135 publications published in the previous five years, researchers systematically reviewed theimpact of machine learning on lockdown exam management systems. Machine learning's impact on each stage of the test process was analyzed and analyzed, from pre-exam preparation through exam administration and assessment. In each step, the Machine learning methods used were recognized and classified as either unsupervised or supervised. Machine learning approaches were studied in depth to examine the most important facets of exams, including authentication, scheduling, proctoring, and cheat or fraud detection. Understanding the function of machine learning in test preparation and its management of the post-exam process requires integrating the primary qualities, such as prediction of at-risk students, adaptive learning, and monitoring of students [8].

G. Latif et al. (2023) provided machine learning to evaluate college students' performance using data collected from their online activities. Researchers pay greater attention to the use of machine learning in education as the number of students at all levels throughout the world rises. Work had to bedone to aid students, teachers/professors, parents, and other stakeholders in providing the assistance students need to succeed in K-12 and higher education settings.The goal of this study was to create a system basedon updated machine learning models to automatically forecast students' performance and then identify pupils at risk since the need and incentive for such systems are well-established. In this analysis, theymake use of the DEEDs data collection. Exciting newcharacteristics were retrieved and used with popular classifiers, including ensemble classifiers. Base learners, such as bagging and boosted, were also integrated with this classifier [9].

Q. liu et al. (2022) explained machine learning and intelligent data recognition in a privacy protectionsystem. Beginning with a discussion of the nature of the privacy threat posed by machine learning, this article goes on to provide a concise summary of the features and techniques used to safeguard user data. Then, a deep-difference privacy protection approach using a CNN was developed to improve the classification accuracy of the different algorithm that employs noise to safeguard confidentiality. Theprivacy budget for each layer of the neural network may be determined using this technique, which flawlessly combines the properties of difference and Gaussian distribution. Finally, the gradient value from the stochastic gradient descent technique is usedto adjust the level of the Gaussian noise and protect the sensitive data. The experimental findings showed that a middle ground could be found between the availability and privacy protection of the training dataset by modifying the parameters of the depthdifferential privacy model based on variations in private information in the data [10].

S. C. Matz et al. (2023) predicted student retention using app-based engagement metrics and socio-demographic factors with machine learning. Academic institutions, financial agencies, and students all face significant difficulties due to student attrition. Researchers in the field of higher education have found that they can use macro-level and micro-level data to predict whether or not a student will drop out of school, thanks to the advent of Big Data and predictive analytics. The present work, however, has largely ignored a vital Meso-level factor of studentperformance known to impact retention: the student's experience at university and their social embeddedness within their cohort [11].

A. Namoun et al. (2021) reviewed data mining and learning analytics as a predictive tool for student outcomes. They analyzed and summarized 62 publications that were pertinent to the topic, with an emphasis on three areas: (1) the methods used toanticipate learning outcomes; (2) the predictiveanalytics models created to foretell student learning; and (3) the most influential aspects in determining learning outcomes. The findings were synthesized and reported using standard procedures for systematic literature reviews, such as PICO and PRISMA. Class rankings and accomplishment grades were the primary indicators of students' success in achieving desired learning outcomes. It was common practice tocategorize students based on their performance using regression and supervised machine learning models. [12].

P. Nikolaidis et al. (2022) introduced determinants of learning progress as a predictor of college dropout rates. The goals of regulations governing higher education were to help students make progress intheir studies and achieve their academic goals. This study presents a research methodology to examine whether at-risk students may be proactively detected via self-evaluation of academic aspects contributingto their learning development, drawing inspiration from Tinto's integration theory and Bean's attrition model. [13].

S. Rajendran et al. (2022) explained machine learning methods for predicting pupils' academic achievement in middle school and high school. This study was one of the first to use Machine Learning Algorithms (MLAs) to predict the academic performance of middle and high school students based on a wide range of demographic (such as age, gender, obesity, average household income, familysize, and parental marital status), institutional (such as gender and academic level), and individual (suchas stress and lifestyle) factors [14].

V. Realinho et al. (2022) predicted whether or not a student will quit school because colleges and universities collect so much information on their students, there was a wealth of opportunity for research, analysis, and oversight. Both high school dropouts and college dropouts impede economic development, employment, competitiveness, and production, with serious consequences for individuals, families, colleges, and communities [15].

J. R. Reidenberg et al. (2018) improved privacy for educational big data. This essay delves into the ethical conflicts that arise between using Big Data forthe sake of education and protecting student privacyin the context of modern learning tools. They proposed that accountability and supervision must be included in learning technologies and thatit is necessary to show that these systems work while protecting users' privacy. They finish by suggesting policy changes that might help accomplish these aims[16].

M. Segura et al. (2022) looked machine learning predicts college dropout. The purpose of this research was to utilize information on students who dropped out of Europe's third-largest traditional university to make predictions about whether or not those students will drop out of school before or after the first semester. They based our estimate on data from all five of the show's main genres. There had been a variety of approaches: Logistic Regression, Support Vector Machines, Decision Trees, Artificial Neural Networks, and a Feature Selection Processwere used to determine which factors were most strongly associated with attrition [17].

D. P. Wick et al. (2022) analyzed the effectiveness of engineering retention programs at a private research institution. The purpose of this study was to verify the accuracy of the binary classifier used to identify incoming freshmen who were not academically prepared and to evaluate the effectiveness of the interventions designed to helpthem [18].

M. Yağcı et al. (2022) predicted pupils' academic achievement via machine learning algorithms throughthe mining of educational data. To betterforecast undergraduate students' final test scores using their midterm exam grades as training data, thisresearch offers a novel model based on machine learning techniques [19].

B. Bayat et al. (2018) reviewed the students' rates of academic performance and contributed variables werebeing evaluated. Academic achievement was a key indicator of a student's potential to succeed academically and progress toward graduation. The success rate in school may be increased by recognizing the contributed component. The purposeof this research was to analyze the elements that contribute to University of Tehran students' academic achievement [20].

S. Boumi et al. (2021) focused on the hidden Markov model, they were able to quantify the effect that students' enrollment habits had on their academic performance. Using a Hidden Markov Model and an advanced multi-period dynamic technique, y classify students' enrollment strategies as either full-time, part-time, or mixed. They then examine and contrast the academic performance results of eachcohort depending on their enrollment tactics, differentiating between first-time freshmen and transfer students [21].

C. Halloran et al. (2021) provided student test scores and the pandemic teaching method: data from thestates of the United States. They evaluated the effect of education modality on standardized test results at the district level. They aggregate information from 12states' 2020-21 school year enrollment records, attendance, and standardized exam scores from the spring of 2021. They find that passed rates have dropped from previous years and that the drop has been more pronounced in districts that have reduced the number of classroom teachers [22].

A. Alyami et al. (2021) looked a t the effects of poortime management on a student's grades. The research method used in this study was a cross-sectional survey. Students at King Abdulaziz University's Department of Diagnostic Radiology Technology participated in the study, which took place between September and November of the corresponding year [23].

3. Factors Influencing the Real-World Application of Machine Learning for Student Success Rate Calculation

The real-world application of machine learning for student success rate calculation in educational institutions is influenced by various factors. These factors can have a significant impact on the effectiveness and adoption of machine learningmodels in this context. Here are some key factors thatinfluence the real-world application of machinelearning for student success rate calculation:

3.1. Data Quality and Availability

The availability of high-quality, relevant data is paramount. Educational institutions musthave access to comprehensive student data, including academic records, attendance, behavioral data, and demographic information. Incomplete or inaccurate data can lead to unreliable predictions and hinder the implementation of machine learningmodels.

3.2. Privacy and Ethical Considerations

Educational data often contains sensitive information about students. Ensuring data privacy and complying with relevant regulations, such as GDPR or FERPA, isessential. Institutions need to establish robust data governance practices and addressethical concerns related to student datausage.

3.3. Feature Engineering

Feature engineering involves selecting and transforming the rightvariables or features for modeling. Domain expertise is crucial in determining which features are relevant for predicting student success. Additionally, feature engineering can be time-consuming and requires collaboration between data scientists and educational experts.

3.4. Model Selection and Evaluation

Choosingthe appropriate machine learning algorithms and models is critical. Different models mayperform differently depending on thecharacteristics of the data. Rigorous model evaluation and validation are necessary toensure that the chosen model is accurate andgeneralizes well to real-world scenarios.

3.5. Scalability

Educational institutions often deal with a large volume of data. Machine learning models must be scalable to handle the increasing amount of data efficiently. Scalability concerns include computational resources and infrastructure.

3.6. Interpretability and Explainability

Understanding why a model makes certain predictions is essential, especially in educational contexts. Transparent and interpretable models are preferred becausethey allow educators to gain insights into thefactors influencing student success and makeinformed decisions.

3.7. User Acceptance

The acceptance ofmachine learning solutions by educators andadministrators is crucial. Users need to trust the model's predictions and feel confident in its ability to improve student outcomes. Providing user-friendly interfaces and clear explanations of model outputs can enhance acceptance.

3.8. Integration with Existing Systems

Machine learning solutions should seamlessly integrate with existing educational systems, such as learningmanagement systems (LMS) and student information systems (SIS). Compatibility and ease of integration can impact the feasibility of adoption.

3.9. Continuous Monitoring and Maintenance

Machine learning models require ongoing monitoring and maintenance to remain effective. Models can become outdated as student populations and educational practices change. Regular updates and retraining are necessary to ensure the accuracy of predictions.

3.10. Resource Constraints

Educational institutions may face budget constraints and limited access to data science expertise. These limitations can affect the feasibility ofimplementing machine learning solutions. Collaboration with external experts or leveraging cloud-based services can help mitigate resource constraints.

In summary, the successful real-world application of machine learning for student success rate calculation in education hinges on data quality, ethical considerations, model selection, scalability, interpretability, user acceptance, system integration, and ongoing maintenance. Comprehensively addressing these factors is essential to harness the potential of machine learning for improving student outcomes and educational experiences.

4. Impact of Machine Learning on Student Achievement and Educational Institutions

Machine learning has the potential to exert aprofound impact on student achievement andeducational institutions in various ways:

4.1. Personalized Learning

Machine learningalgorithms can analyze individual student data, such as performance, learning preferences, and engagement levels, to create customized learning paths and content. This tailoring of educationalexperiences can enhance student engagement and achievement by addressing each student's unique needs.

4.2. Early Intervention

Machine learningmodels can identify at-risk students byrecognizing patterns in their academic performance and behavior. This early warning system allows educators to intervene promptly and provide additionalsupport, reducing dropout rates and improving overall student success.

4.3. Adaptive Learning Platforms

Educationalinstitutions can implement adaptive learning platforms powered by machine learning to dynamically adjust the difficulty and pace oflessons based on individual student progress.This ensures that students neither struggle with overly challenging material nor becomebored with content that is too easy.

4.4. Predictive Analytics

Machine learning canforecast various educational outcomes,including student performance, enrollmenttrends, and resource needs. Institutions can use these insights to optimize resourceallocation and strategic planning.

4.5. Automated Grading and Assessment

Machine learning can automate grading and assessment processes, saving educators time and providing students with quicker feedback. This can lead to more efficient useof instructional time and more detailedfeedback to aid in learning.

4.6. Content Recommendation

Similar toplatforms like Netflix and Amazon, machine learning can suggest relevant educational materials, courses, or study resources to students.

This personalized contentrecommendation can enhance learning experiences and encourage the exploration of related topics.

4.7. Enhanced Teaching Tools

Machine learning can provide educators with tools forcreating interactive and engaging content, identifying areas where students struggle, and suggesting instructional improvements. These tools can help instructors deliver more effective lessons.

4.8. Language Processing and Translation

Machine learning-driven language processing tools can assist non-native speakers in understanding and communicating in their chosen language of study. Translation and language correction features can improve language learning.

4.9. Administrative Efficiency

Machine learning can automate administrative tasks such as scheduling, course planning, and resource allocation, freeing up institutional resources and personnel for more strategic roles.

4.10. Learning Analytics

Educational institutions can utilize machine learning to analyze large volumes of data related to student performance, behavior, andengagement. These insights can inform decision-making processes to improve curriculum design and teaching strategies.

4.11. Support for Special Needs

Machine learning applications can be tailored to assiststudents with special needs. For example,speech recognition and text-to-speech tools can aid students with speech or reading difficulties.

4.12. Research Advancements

Machine learning enables educational researchers to conduct more sophisticated data analysis andmodeling, leading to deeper insights into learning processes and the development of evidence-based teaching methods.

Despite these positive impacts, it's important to address potential challenges related to dataprivacy, algorithmic bias, and the need for teacher training in using machine learning tools effectively.

Ethical considerations must alsoguide the responsible implementation of machinelearning in education to ensure equitable outcomes for all students.

4.13. Scope of research

Comprehensive analysis of factors influencingthe real-world application of machine learning for student success rate calculation and their impacts on student achievement and educational institutions is extensive and multifaceted. Such an analysis would encompass various dimensionsand involve a combination of qualitative and quantitative research methods. Here is an overview of the potential scope:

4.14. Data Analysis and Quality

- Examination of the quality, availability,and completeness of educational data.
- Assessment of data privacy and securityconsiderations.
- Analysis of historical data to identifytrends and patterns.

4.14.1. Model Development and Evaluation

- Selection of machine learning algorithms and techniques suitable for student success prediction.
- Evaluation of model performance, including accuracy, precision, recall, and F1-score.
- Comparison of different models and approaches.

4.14.2. Feature Engineering

- Identification of relevant features and variables for modeling.
- Exploration of domain-specific factors influencing student success.
- Dimensionality reduction techniques and feature selection strategies.

4.14.3. Ethical Considerations

- Evaluation of potential bias in the data and its impact on model fairness.
- Assessment of ethical implications related to data collection and usage.
- Strategies for mitigating bias and ensuring fairness in predictions.

4.14.4. Privacy and Compliance

- Review of data protection regulations (e.g., GDPR, FERPA) and their implications.
- Development of data governance policies and procedures.
- Strategies for obtaining informed consent for data usage.

4.14.5. Integration and Scalability

- Examination of integration challenges with existing educational systems.
- Assessment of the scalability of machine learning models for large institutions.
- Consideration of computational resources and infrastructure requirements.

4.14.6. User Acceptance and Engagement

- Surveys and interviews to gauge the acceptance and satisfaction of educators, students, and administrators.
- Feedback mechanisms for continuous improvement based on user input.

4.14.7. Predictive Analytics and Early Intervention

- Monitoring the effectiveness of machine learning in predicting student success.
- Assessment of early intervention strategies and their impact on student outcomes.
- Case studies and success stories of institutions implementing predictive analytics.

4.14.8. Administrative Efficiency

- Analysis of how machine learning streamlines administrative processes.
- Quantification of time and resource savings achieved through automation.
- Cost-benefit analysis of machine learning adoption in educational institutions.

4.14.9. Student Achievement and Outcomes

- Investigation of the impact of personalized learning on student achievement.
- Evaluation of whether interventions based on machine learning predictions lead to improved outcomes.
- Long-term studies to assess the effects on graduation rates and career success.

4.14.10. Educational Research and Continuous Improvement

- Utilization of machine learning for educational research, including curriculum design and pedagogical improvements.
- Measurement of the contribution of machine learning to institutional excellence.

4.14.11. Challenges and Limitations

- Identification and analysis of challenges faced during implementation.
- Evaluation of limitations, including the need for data infrastructure and expertise.

4.14.12. Best Practices and Recommendations

- Development of best practices for implementing machine learning in education.
- Recommendations for educational institutions considering adopting machine learning for student success initiatives.

5. Discussion

K. Al Mayahi presented Machine Learning to Improve Academic Performance in the Classroom in 2020. They included past student data and test scores in their algorithm. The model's high rate ofaccuracy helped convince others to use it [1].Machine learning for forecasting student performancewas the topic of a presentation by M. Anusha in2019. The goal of their research was to identify students at risk of failing to meet expected standards. Using machine learning to monitor and predict students' development and performance would significantly improve their system [2]. In 2022, C. J. Arizmendi used data mining techniques to make predictions based on student behavior from their digital classroom recordings. As a way to this end, they provide feature mapping and analysis of LMS data. They concluded with an actual example, emphasizing essential components, and testing model performance across a variety of model settings, to demonstrate the feasibility of making reliable predictions using LMS data [3]. R. B. Saglam elaborated on how humans see, classify, and acquire self-awareness in 2022. By analyzing how privatedata has evolved and been understood differently by these key players, this research contributes to ourunderstanding of the fundamental disconnects that exist and the societal ramifications of emerging technologies [4]. T. Cardona presented a data-mining and machine-learning-based approach for student retention in higher education in 2023 [5]. K. Hamann investigated student persistence and performance in distance learning political science programs in 2021. They looked at the effects that students' preferences for course delivery modes have on their grades. Students' academic performance tends to decline as their dependence on online courses grows [6].

B. Holicza studied machine learning methods in 2023 to forecast and evaluate students' academic achievement in both virtual and physical classrooms. Predictions of student performance are made and evaluated using a variety of machine-learning techniques in this research. These techniques include the support vector machine with varying kernels, the decision tree, the random forest, and the k-nearest neighbor's algorithm. They concluded that habits like sleeping, studying, and using electronic media all contribute to academic success [7]. S. Kaddoura discussed machine learning strategies for use in digital assessment and instructionin 2022. Test authentication, test scheduling, proctoring, and cheat/fraud detection were some ofthe most explored machine learning techniques in thisstudy. Machine learning's core properties, such as at-risk student prediction, adaptive learning, and monitoring, must be integrated to comprehend its role in test preparation and its management of the post-exam process [8]. In 2023,

G. Latif pioneered the use of machine learning to assess pupils' academic progress based on their digital footprint. As the number of students at alllevels throughout the globe continues to climb, researchers are paying more attention to the use of machine learning in the classroom. To help kids thrive in elementary, secondary, and tertiary education, work had to be done to aid students, teachers/professors, parents, and other stakeholders. We recovered exciting new features and applied themto well-known classifiers, such as ensemble classifiers. These classifiers were also used with base learners like bagging and boosted [9]. Q. liu proposeda privacy protection solution based on machine learning and intelligent data identification in 2022. After introducing the concept of the privacy hazard presented by machine learning, this article provides a high-level overview of the safeguarding features and strategies already in use. Then, a CNN-based deep difference privacy protection method was created to boost the different algorithms' classification precision, which makes use of noise to conceal sensitive information. The results of the experiments demonstrated that by adjusting the parameters of the depth differential privacy model based on fluctuations in private information in the data, a compromise could be established between the availability and privacy protection of the training dataset [10].

S. C. Matz used app-based engagement indicators and socio-demographic characteristics withmachine learning to estimate student retention in 2020 [11]. The predictive power of data mining and learning analytics was examined by A. Namoun in a 2021 article. Standard methods for systematic literature reviews, such as PICO and PRISMA, were used to synthesize and publish the results. Students' progress toward learning objectives was measured primarily by their placement in the class and their performance marks. Traditional methods of student classification included regression and supervisedmachine-learning algorithms [12].

Limitations of conventional research were issues of performance due to huge datasets. Moreover there remained an accuracy issue. To resolve the issue of performance there remained a requirement of a data compression mechanism. on the other hand, to improve the accuracy there was a requirement for an advanced deep learning approach that should be capable of resolving the issue of accuracy.

6. Conclusion

A comprehensive analysis of these factors would involve collaboration between data scientists, educators, administrators, and researchers. It would also require a longitudinal approach toassess the long-term impacts of machine learningon student achievement and educational institutions. Such an analysis can provide valuable insights into the benefits and challengesof adopting machine learning in education and guide institutions in making informed decisions to improve student outcomes.

Compliance with ethical standards

Acknowledgments

I would like to extend my gratitude to Ms. Aditi Misra, Director-principal of our school, and my teachers, mentors, and parents for motivating me and guiding me in this project.

This research was done under the mentorship and esteemed guidance of Prof Jamie Fairclough with articles collected for the Stanford Pre-Collegiate Summer Program course entitled “Introduction to Machine Learning” conducted at Stanford University, USA.

Disclosure of conflict of interest

No conflict of interest is to be disclosed.

References

[1] K. Al Mayahi and D. M. Al-Bahri, “Machine Learning Based Predicting Student Academic Success,” *Int. Congr. Ultra Mod. Telecommun. Control Syst. Work.*, vol. 2020-October, no. February, pp. 264–268, 2020, doi: 10.1109/ICUMT51630.2020.9222435.

[2] M. Anusha, K. Karthik, P. Padmini Rani, and V.Srikanth, *Prediction of student performance using machine learning*, vol. 8, no. 6. Atlantis Press International BV, 2019. doi: 10.35940/ijeat.E7520.088619.

[3] C. J. Arizmendi *et al.*, “Predicting student outcomes using digital logs of learningbehaviors: Review, current standards, and suggestions for future work,” *Behav. Res. Methods*, no. 0123456789, 2022, doi: 10.3758/s13428-022-01939-9.

[4] R. Belen Saglam, J. R. C. Nurse, and D. Hodges, “Personal information: Perceptions,types and evolution,” *J. Inf. Secure. Appl.*, vol. 66, no. March, p. 103163, 2022, doi: 10.1016/j.jisa.2022.103163.

[5] T. Cardona, E. A. Cudney, R. Hoerl, and J. Snyder, "Data Mining and Machine Learning Retention Models in Higher Education," *J. Coll.Student Regent. Res. Theory Pract.*, vol. 25, no. 1, pp. 51–75, 2023, doi: 10.1177/1521025120964920.

[6] K. Hamann, R. A. Glazier, B. M. Wilson, and P.H. Pollock, "Online teaching, student success, and retention in political science courses," *Eur. Polit. Sci.*, vol. 20, no. 3, pp. 427–439, 2021,doi: 10.1057/s41304-020-00282-x.

[7] B. Holicza and A. Kiss, "Predicting and Comparing Students' Online and OfflineAcademic Performance Using Machine Learning Algorithms," *Behav. Sci. (Basel).*, vol. 13, no. 4, 2023, doi: 10.3390/bs13040289.

[8] S. Kaddoura, D. E. Popescu, and J. D. Hemanth,"A systematic review on machine learning models for online learning and examination systems," *PeerJ Comput. Sci.*, vol. 8, no. Ml,pp. 1–32, 2022, doi: 10.7717/PEERJ-CS.986.

[9] G. Latif, S. E. Abdelhamid, K. S. Fawagreh, G. Ben Brahim, and R. Alghazo, "Machine Learning in Higher Education:Students’ Performance Assessmentconsidering Online Activity Logs," *IEEE Access*, vol. PP, p. 1, 2023, doi: 10.1109/ACCESS.2023.3287972.

[10] Q. Liu, "Privacy Protection Technology Based on Machine Learning and Intelligent DataRecognition," *Secur. Commun. Networks*, vol. 2022, 2022, doi: 10.1155/2022/1598826.

[11] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons,and C. Stachl, "Using machine learning to predict student retention from socio-demographic characteristics and app-basedengagement metrics," *Sci. Rep.*, vol. 13, no. 1, pp. 1–16, 2023, doi: 10.1038/s41598-023- 32484-w.

[12] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematicliterature review," *Appl. Sci.*, vol. 11, no. 1, pp. 1–28, 2021, doi: 10.3390/app11010237.

[13] P. Nikolaidis, M. Ismail, L. Shuib, S. Khan, and G. Dhiman, "Predicting Student Attrition in Higher Education through the Determinants of Learning Progress: A Structural EquationModelling Approach," *Sustain.*, vol. 14, no. 20, 2022, doi: 10.3390/su142013584.

[14] S. Rajendran, S. Chamundeswari, and A. A. Sinha, "Predicting the academic performance of middle- and high-school students using machinelearning algorithms," *Soc. Sci. Humanit. Open*, vol. 6, no. 1, p. 100357, 2022, doi: 10.1016/j.ssaho.2022.100357.

[15] V. Realinho, J. Machado, L. Baptista, and M. V.Martins, "Predicting Student Dropout and Academic Success," *Data*, vol. 7, no. 11, 2022, doi: 10.3390/data7110146.

[16] J. R. Reidenberg and F. Schaub, "Achieving big data privacy in education," *Theory Res. Educ.*, vol. 16, no. 3, pp. 263–279, 2018, doi: 10.1177/1477878518805308.

[17] M. Segura, J. Mello, and A. Hernández, "Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role?," *Mathematics*, vol. 10, no. 18, pp. 1–20, 2022, doi: 10.3390/math10183359.

[18] D. P. Wick *et al.*, "Measuring the impact ofstudent success retention initiatives for engineering students at a private research university," *Front. Educ.*, vol. 7, 2022, doi: 10.3389/feduc.2022.1087064.

[19] M. Yağcı, "Educational data mining: prediction of students' academic performance usingmachine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, 2022, doi:10.1186/s40561-022-00192-z.

[20] S. Mohebi, M. Parham, G. Sharifirad, and Z. Gharlipour, "Social Support and Self-Care Behavior Study," no. January, pp. 1–6, 2018, doi: 10.4103/jehp.jehp.

[21] S. Boumi and A. E. Vela, "Quantifying the impact of student enrollment patterns on academic success using a hidden Markov model," Appl. Sci., vol. 11, no. 14, 2021, doi: 10.3390/app11146453.

[22] C. Halloran, R. Jack, J. C. Okun, and E. Oster, "Pandemic Schooling Mode And Student Test Scores: Evidence From Us States," NBER Work. Pap., pp. 1–28, 2021.

[23] A. Alyami, A. Abdulwahed, A. Azhar, A. Binsaddik, and S. M. Bafaraj, "Impact of Time- Management on the Student's AcademicPerformance: A Cross-Sectional Study," Creat. Educ., vol. 12, no. 03, pp. 471–485, 2021, doi 10.4236/ce.2021.123033.