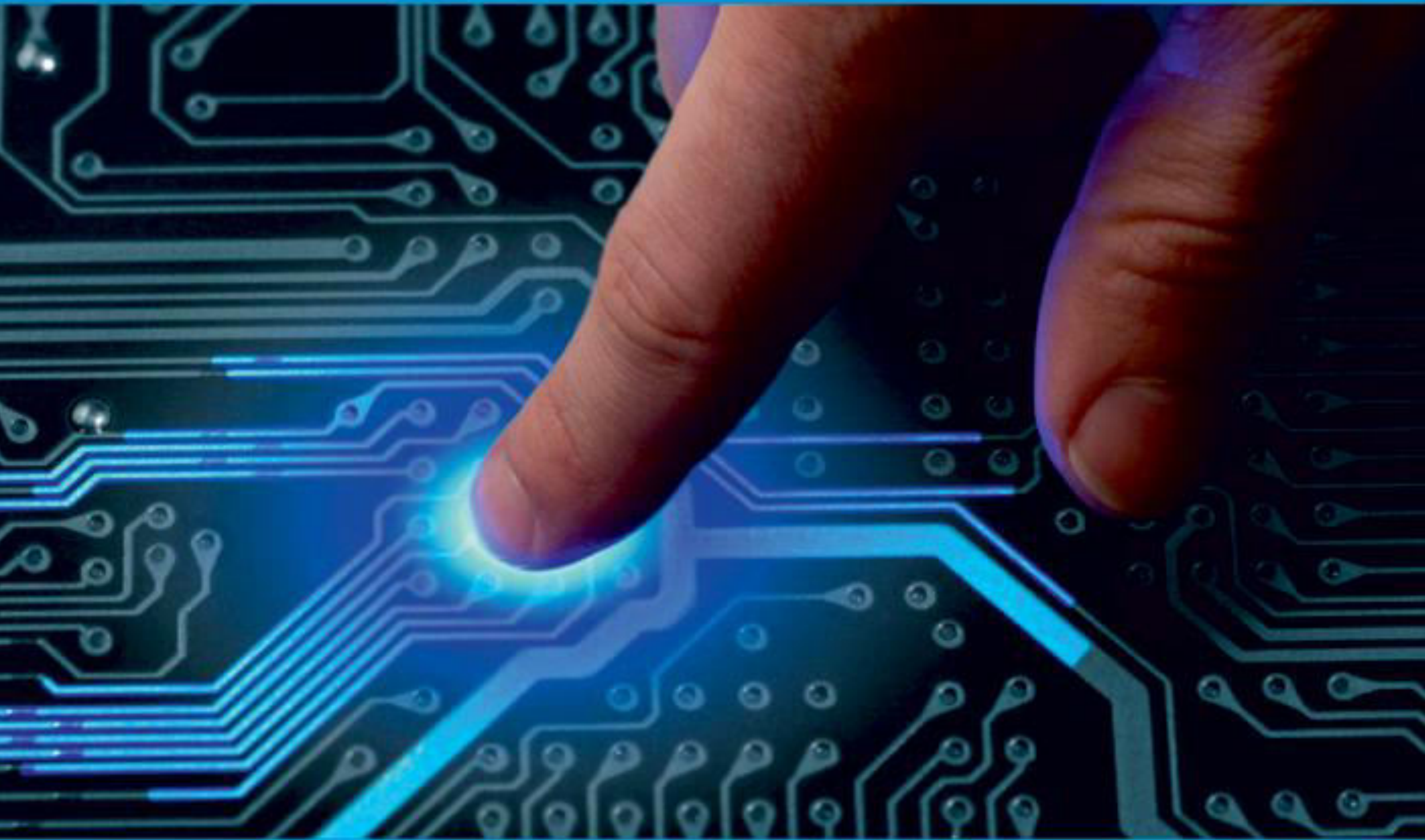
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 11, November 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || A Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 11, November 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111044 |

Early Detection of Students at Risk Using Machine Learning to Minimize Failure in Academics

Fija N. Chaugule, Komal K. Hule, Rutika A. Khandagale, Dr. Mrs. Shubhangi Gunjal.

UG Student, Dept. of C.S, Jaihind College of Engineering, Kuran, Pune, Maharashtra, India

Assistant Professor, Dept. of C.S. Jaihind College of Engineering, Kuran, Pune, Maharashtra, India

ABSTRACT: This research strongly emphasizes the important role of machine learning in education, especially when it comes to predicting and improving how well students do in school. It believes in the old saying that ”prevention is better than a cure” and wants to help teachers by giving them tools to support their students at the right time.One big thing this research does is find out what things affect how well students do on their final exams and suggests which classes might be best for them. This helps teachers make smart choices for their students. The research also talks about the limits of using people to keep an eye on students and really wants to use smart computers to help teachers understand what students need.The main goal of all this is to help students do better in school by figuring out how they might do in the future and giving suggestions on what to do next. So, in short, this study is all about exploring how machine learning, education, and support for students can work together. It’s about dealing with the problems in education and saying that it’s really important to help students early on.

KEYWORDS: Machine Learning, Students Performance, Early Detection, Educational Technology, Classification, Hybrid model.

I. INTRODUCTION

Machine learning is making a big impact in education, offering promising opportunities for improving teaching, learning, research, and decision-making. It essentially involves making machines mimic human reasoning and behavior. Some machine learning algorithms, like Na¨ıve Bayes and decision trees, have the ability to predict and are being used to enhance the learning experience. These algorithms learn from data and can act in ways that resemble human actions. They’re particularly useful in understanding each student’s strengths and weaknesses and why they might struggle in their studies.

Forecasting learning outcomes using machine learning can help educational institutions understand how students learn, Machine learning can reduce forecast inaccuracies that have plagued forecasting systems and the potential for poor decision-making and system failures. These models can also monitor students’ progress and provide recommendations for personalized teaching and support. However, while there’s a lot of promise in using machine learning for academic forecasting, it’s important to acknowledge its limitations. This study delves into the benefits and shortcomings of machine learning in this context and aims to provide valuable insights for forecasters and education authorities

Student retention and success in higher education have been persistent challenges. Retention rates have remained low, with a significant percentage of students leaving after their first year. Academic success is a key factor in student retention, and the best predictor of persistence. Identifying at-risk students early in the semester is crucial to increasing academic success. Predictive modeling techniques can serve as early warning systems, identifying students in need of support and intervention.Current early warning systems, while promising, have limitations. They often use general prediction models that may not account for the complexity of various courses. They also rely heavily on Course Management System (CMS) data, which may not be suitable for face-to-face courses. Additionally, these systems may not effectively identify at-risk students who need the most accurate predictions. Furthermore, early warning systems are often based on traditional summative. score-based grading systems, and as more institutions shift to standards-based grading, these systems need to adapt.

while promising, have limitations. They often use general prediction models that may not account for the complexity of various courses. They also rely heavily on Course Management System (CMS) data, which may not be suitable for face-to-face courses. Additionally, these systems may not effectively identify at-risk students who

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || A Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 11, November 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111044 |

need the most accurate predictions. Furthermore, early warning systems are often based on traditional summative.

II. RELATED WORK

Predictive Modeling with Academic Data: Many studies have focused on utilizing academic data, such as grades, attendance records, and course engagement, to build predictive models for identifying students at risk. Machine learning algorithms, including decision trees, logistic regression, and neural networks, have been applied to analyze historical data and predict future academic performance.

Feature Selection and Engineering: Researchers have explored different features that contribute to predicting student success or failure. This includes not only academic data but also non-academic factors such as socio-economic status, attendance, and participation in extracurricular activities. Feature selection and engineering techniques have been employed to identify the most relevant factors.

Natural Language Processing (NLP) for Text Analysis: Some studies have leveraged NLP techniques to analyze textual data, such as essays, forum posts, or communication between students and teachers. Sentiment analysis and topic modeling have been used to gain insights into students' emotional well-being and engagement, contributing to early detection efforts.

Intervention Strategies: Beyond prediction, research has delved into effective intervention strategies for at-risk students. This includes personalized learning plans, mentoring programs, and targeted support systems. Machine learning has been employed to recommend specific interventions based on individual student profiles.

Early Warning Systems in Educational Technology: Educational technology platforms have incorporated early warning systems that use machine learning algorithms to monitor students' progress and provide timely alerts to educators. These systems often analyze various data sources to generate actionable insights.

Longitudinal Studies and Retrospective Analyses: Longitudinal studies have been conducted to understand the factors leading to academic success or failure over an extended period. Retrospective analyses of past academic data have provided valuable insights into patterns and trends that can inform the development of more accurate predictive models.

Ethical Considerations: With the increasing use of machine learning in education, there is a growing focus on ethical considerations. Researchers have explored issues related to bias in algorithms, privacy concerns, and the responsible use of predictive analytics in the educational context.

Comparative Studies and Benchmarking: Comparative studies have been conducted to evaluate the effectiveness of different machine learning algorithms and models in predicting student outcomes. Benchmarking efforts aim to establish standard practices and metrics for assessing the performance of predictive models.

Integration with Learning Management Systems (LMS): Some research has focused on integrating predictive analytics directly into Learning Management Systems, providing real-time feedback to educators and students. This seamless integration facilitates more immediate and targeted interventions.

Collaborative Research and Open Source Initiatives: Collaborative research efforts and open-source initiatives have emerged to encourage the sharing of datasets, methodologies, and code. This collaborative approach fosters the development of more robust and generalizable models for early detection.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || A Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 11, November 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111044 |

III. ALGORITHM

Logistic Regression:
Logistic regression is a statistical method used for predicting the probability of an event happening, usually a binary outcome (like yes/no, success/failure). It's named "logistic" because it uses the logistic function to model the relationship between the independent variables and the probability of the event occurring. Essentially, it's a tool for classification tasks.

Decision Tree:
A decision tree is a supervised machine learning algorithm used for both classification and regression tasks. It works by recursively partitioning the input space into regions and assigning a label or predicting a target value in each region. The decision tree structure is hierarchical and resembles an upside-down tree, where each node represents a decision based on the value of a specific feature. The leaves of the tree contain the final output, either a class label in the case of classification or a numerical value for regression.

Naive Bayes:
Naive Bayes is a classification algorithm that is based on Bayes' theorem with an assumption of independence among features. Despite its simplicity and the "naive" assumption, it often performs surprisingly well in practice, especially for text classification tasks.

Random Forest:
Random forest is an ensemble learning algorithm that uses a collection of decision trees to make predictions. Random forest as a group of decision trees working together. Each tree in the forest makes its own prediction, and then the "forest" combines all these predictions to give you a more accurate and reliable answer. It's like asking multiple friends for advice on a decision – each friend has their perspective, and by considering all of them, you can make a better decision. The "random" part comes from using different subsets of the data and features for each tree, adding diversity to the group.

AdaBoost:
AdaBoost (Adaptive Boosting) is an ensemble learning algorithm that is used for classification and regression tasks. It is particularly effective in boosting the performance of weak learners (classifiers that perform slightly better than random chance) to create a strong classifier. AdaBoost was introduced by Yoav Freund and Robert Schapire in 1996.

K-Nearest Neighbours:
The k-Nearest Neighbors (k-NN) algorithm is a supervised machine learning algorithm used for classification and regression tasks. It is a simple and intuitive algorithm that classifies a data point based on the majority class (for classification) or averages the values (for regression) of its k-nearest neighbors in the feature space.

Support Vector Machine:
Support Vector Machines (SVM) is a supervised machine learning algorithm that can be used for classification or regression tasks. SVM is particularly effective in high-dimensional spaces and is well-suited for both linear and nonlinear problems. The primary objective of SVM is to find a hyperplane that best separates the data points of one class from another while maximizing the margin between the classes.

IV. PSEUDOCODE

Step-1 **Data Collection:**

The study begins with the collection of in-semester performance data.

Step-2 **Data Preprocessing:**

This stage involves thorough data cleaning and preparation to ensure that the data is free from errors, inconsistencies, and missing values.

Step-3 **Feature Selection:**

To enhance the generalizability and accuracy of the predictive models, a feature selection method is utilized.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || A Monthly Peer Reviewed & Referred Journal |

|| Volume 11, Issue 11, November 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1111044 |

Step-4 **Model Selection:**

The choice of predictive modeling methods is a critical decision.

Step -5 **Training and Testing:**

To evaluate the models, the dataset is divided into a training set for model development and a testing set for model evaluation.

Step-6 **Performance Metrics:**

The study defines specific performance metrics that are used to assess the models' effectiveness in identifying at-risk students.

V. CONCLUSION AND FUTURE WORK

To sum it up, using machine learning to identify students at risk is a big step forward in education. A review of research shows that these methods work well in predicting how students will do, whether we look at their past grades or a different system. This is a good way to step in early and help students who are having a tough time, which leads to better results in education. Combining Educational Data Mining and machine learning is a valuable tool for schools to find students who might need help, improve how well they do in their studies,and give them the right support to succeed.

REFERENCES

1. M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," IEEE Access, vol. 9, pp. 7519–7539, 2021.
2. R. Z. Pek, S. T. Ozyer , T. Elhage , T. Ozyer AND R. Alhajj ¨ "The Role Of Machine Learning in Identifying Students At Risk And Minimizing Failures" The Scientific and Technological Research Institution of Turkey (TUBITAK) under Program Grant 2209-A,IEEE Xplore/Access,January 2023
3. G. Sujatha, S. Sindhu, and P. Savaridassan, "Predicting students performance using personalized analytics," Int. J. Pure Appl. Math., vol. 119, no. 12, pp. 229–238, 2018.Roy, Abhimanyu, et al: Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.
4. R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, "Detecting at risk students with early interventions using machine learning techniques," IEEE Access, vol. 7, pp. 149464–149478, 2019.
5. A. Behr, M. Giese, H. D. Teguim, and K. Theune, "Early prediction of university dropouts—A random forest approach," Jahrb¨ucher Nation al¨okonomie Statistik, vol. 240, no. 6, pp. 743–789, Feb. 2020.
6. S. Lee and J. Y. Chung, "The machine learning-based dropout early warning system for improving the performance of dropout prediction," Appl. Sci., vol. 9, no. 15, p. 3093, Jul. 2019.
7. E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Ara´ujo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," Comput. Hum. Behav., vol. 73, pp. 247–256, Aug. 2017.
8. M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," Artif. Intel. Rev., vol. 52, no. 1, pp. 381–407, Jun. 2019.
9. N. Mduma, K. Kalegele, and D. Machuve, "Machine learning approach for reducing students' dropout rates," Int. J. Adv. Comput. Res., vol. 9, no. 42, pp. 156–169, May 2019.
10. J. Y. Chung and S. Lee, "Dropout early warning systems for high school students using machine learning," Children Youth Services Rev., vol. 96, pp. 346–353, Jan. 2019