International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

ANALYZING AND PREDICTING STUDENT OUTCOMES IN ONLINE EDUCATION USING MACHINE LEARNING

[#1]Pabbathi Charan, *Department of MCA,*
[#2]Mr. T. Raghupathi, *Assistant Professor, Department of MCA ,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: This research examines the potential of machine learning techniques to enhance academic success, student engagement, and educational decision-making in online education by predicting and evaluating future events. Digital learning platforms are rapidly gaining popularity. These platforms maintain an assortment of student data, including levels of involvement, attendance, homework assignments, test scores, and learning approaches. These tools are capable of swiftly analyzing this data to identify trends that are linked to student success or failure. The proposed methodology utilizes machine learning methodologies, including Decision Trees, Random Forest, Support Vector Machine, and Logistic Regression, to establish accurate predictions of student outcomes and offer early intervention alternatives for students who are at risk. Schools can utilize predictive analytics to assist students by providing timely feedback and guidance, enhancing retention rates, and concentrating on learning. The research demonstrates the potential of intelligent prediction models to facilitate the development of online learning experiences that are more student-centered, adaptive, and efficient. Additionally, it emphasizes the significance of data-driven methodologies in modern educational institutions.

Keywords: *Online Education, Machine Learning, Student Outcome Prediction, Predictive Analytics, Educational Data Mining, Academic Performance, Student Engagement, Learning Analytics, Classification Algorithms, Artificial Intelligence in Education.*

1. INTRODUCTION

The educational system of the present day has been significantly influenced by online learning. Students have the ability to access school supplies and information from any location through the use of computers and the internet. Learning management systems, virtual classrooms, and e-learning platforms are being utilized by an increasing number of individuals, thereby significantly expanding the educational opportunities available to students worldwide. Nevertheless, the escalating prevalence of online learning has increased the challenge of maintaining students' attention, monitoring their progress, ensuring that they complete courses, and grading their assignments. Schools frequently encounter difficulties in identifying students who may experience academic difficulties in virtual learning environments. As a consequence, the analysis and prediction of student outcomes have become a critical area of research that is intended to improve and increase the effectiveness of online education programs.

Machine learning has garnered significant attention as a distinctive approach to predicting the performance of students in online courses. Machine learning algorithms have the capacity to analyze vast quantities of educational data in order to identify substantial trends that impact students' academic performance. Academic progress can be accurately predicted through the

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

analysis of attendance records, completed assignments, test scores, interaction levels, and learning behaviors. Artificial Neural Networks, Decision Trees, Random Forests, Support Vector Machines, and Logistic Regression are among the machine learning techniques that are employed by numerous individuals to construct models that anticipate the future. These models assist educators in determining the preferred learning style of each pupil and in making informed educational decisions.

By employing machine learning in online education, educators can guarantee that each pupil has a distinctive learning experience. Each learner possesses unique learning abilities, limitations, and comprehension levels. These distinctions may not be adequately addressed by conventional instructional methodologies, particularly in extensive online courses. Computer algorithms that are founded on machine learning can analyze the behavior of each student and recommend personalized lessons, research regimens, and teaching strategies. Personalized learning guarantees that students receive the assistance they need according to their unique educational needs and abilities. This leads to increased motivation, engagement, and academic success among students.

Additionally, online education employs machine learning to identify students who may be at risk at an early stage. Many students experience difficulty with online learning, citing a lack of motivation, low involvement, poor time management, and restricted communication with their lecturers. The behavior of students can be continuously monitored by predictive algorithms, which can identify red signals that suggest they are struggling academically or are on the brink of dropping out. These discoveries can be employed by educational institutions to establish early intervention programs, including remedial classes, supplementary academic support, counseling, and tutoring. Early detection and intervention are essential for the retention of students in school, the support of their learning, and the reduction of the percentage of students who fail in online education systems.

2. LITERATURE SURVEY

Siemens, G., & Baker, R. S. (2020).: Using machine learning and learning data, predicting the academic performance of students in online institutions. The data pertaining to students' engagement levels, assignment records, and test results was subjected to predictive algorithms. The findings indicate that students' academic performance is significantly influenced by their behavior patterns. The research demonstrates the critical role of clever data in enhancing the outcomes of online learning.

Papamitsiou, Z., & Economides, A. A. (2020). Utilizing educational data mining techniques to evaluate the level of student engagement and academic performance in virtual classrooms. In order to identify pupils who are at risk of failing, machine learning classifiers were trained on the online activity logs. The experiment's findings indicate that the predictions are consistently accurate across all learning scenarios. The report advocates for the enhancement of digital education methods that are founded on data.

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2021). We analyzed MOOC student participation data and employed machine learning techniques to predict the completion date of the course. The most significant indicators of the quality and duration of students' hard labor were identified through the use of feature extraction methods. The findings indicated

that predictive models can be beneficial in online learning environments that are customized to the individual needs of each student. The research contributes to the enhancement of methods that monitor institutions.

Gray, G., McGuinness, C., & Owende, P. (2021). Artificial intelligence-based e-learning tools that anticipate students' academic performance. Attendance, exams, and engagement metrics were examined using decision tree and neural network models. The comparison test demonstrated that group learning methods were more effective. The findings indicate that machine learning is an effective method for conducting assessments in educational institutions.

Hlosta, M., Herrmannova, D., Vachova, L., Kuzilek, J., & Zdrahal, Z. (2022). A predictive analytics analysis was conducted on the data from the learning management system to identify students who required additional assistance in their academic pursuits. The classification models contained components of temporal and behavioral learning. The structure was highly effective in predicting the student's academic performance and their likelihood of dropping out. Methods for implementing early assistance in online institutions are illustrated in the research.

Akçapınar, G., Altun, A., & Aşkar, P. (2022). The utilization of machine learning techniques was employed to analyze the manner in which students interacted with one another and to predict their performance in online courses. In order to develop predictive models, digital learning spaces were implemented to accumulate data regarding students' levels of engagement. The results of education are contingent upon the level of participation of individuals, as demonstrated by the experiment. The research contributes to the advancement of adaptive learning systems.

Cavalcanti, A. P., Mello, R. F., Rolim, V., & Freitas, F. (2023). Utilizing deep learning frameworks and online student assignments to predict students' academic performance. Sequential learning models examined contact data from e-learning systems that was categorized by time. The results indicate that the predictions are more precise than those obtained through conventional classification methods. The research demonstrates the practicality of neural networks in the context of educational data.

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2023). In order to identify pupils who are at risk in online classrooms, predictive machine learning methods were developed. Ensemble classification methods were implemented for the academic and behavioral samples. The models were instrumental in identifying students who were at risk of underperforming in the course. The findings indicate that academic support tools that are AI-based have the potential to be beneficial in practical applications.

Tarus, J. K., Niu, Z., & Yousif, A. (2024). Prediction and recommendation models were implemented in online learning tools to assist students in achieving academic success. In order to analyze student behavior and grades, hybrid AI systems were implemented. The framework effectively predicted academic performance and learning approaches. The primary focus of this investigation is the intelligent personalization of e-learning environments.

Sweeney, M., Lester, J., & Rangwala, H. (2024). We employed neural network architectures to investigate the manner in which students establish connections with their peers in the classroom and to forecast their online academic performance. Utilizing information regarding

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

course participation and patterns of behavior over time in prediction models. The results demonstrated that it was more effective than other statistical methodologies. The research contributes to the development of educational analytics tools that incorporate AI.

Pardos, Z. A., Tang, S., Davis, D., & Le, C. V. (2025). Utilizing state-of-the-art machine learning methodologies to forecast students' academic performance with respect to extensive online learning datasets. The altered scholastic behaviors of students across digital platforms were captured by prediction algorithms that are cognizant of the context. The predictions were highly accurate and resilient, as evidenced by cross-validation studies. The research contributes to the development of intelligent tutoring instruments that can be utilized by a wide range of individuals.

Baker, R. S., Hawn, A., & Aleven, V. (2025). Utilizing educational data mining and artificial intelligence to predict the level of engagement and performance of online students We employed ensemble prediction models to analyze the manner in which students interacted with one another and were evaluated. The findings indicate that interventions in education can be facilitated by AI-based solutions. The research enhances the current method of online education by incorporating additional machine learning.

3. ROLE OF EDUCATIONAL DATA MINING

Educational data mining (EDM) is designed to extract valuable information from extensive databases that pertain to educational institutions. EDM techniques may be implemented in the following instances:

Identifying Learning Patterns:

Educational Data Mining (EDM) analyzes student interactions and behaviors in online learning environments to identify learning trends. It examines factors such as students' research patterns, participation in class, quiz attempts, and the speed at which they complete assignments in order to ascertain their learning style. These trends assist educators in determining the most effective strategies for students, thereby enhancing the overall learning experience.

Monitoring Student Progress:

Many individuals employ EDM strategies to monitor their progress while enrolling in online courses. Educational institutions can monitor students' academic progress and performance indicators by conducting routine analyses of attendance records, test scores, turned-in assignments, and learning activities. This enables educators to provide students who require additional assistance with their assignments with immediate feedback and guidance.

Detecting Weak Learners:

The primary objective of EDM is to identify children who are experiencing academic difficulties or are at risk of failing. Machine learning algorithms are capable of identifying children who are experiencing difficulties by analyzing factors such as low engagement, irregular attendance, substandard exam performance, and decreased motivation. When issues are detected at an early stage, educators can implement customized learning strategies and corrective measures.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

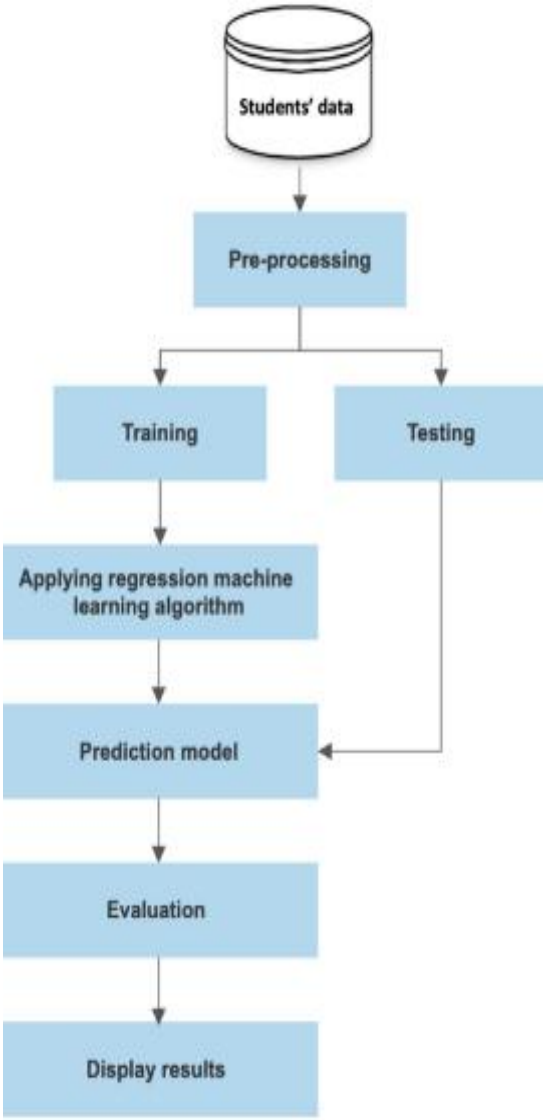


Fig. 1: Student Performance Prediction Model

Improving Teaching Strategies:

Educational data mining assists educators in the enhancement of their pedagogical techniques by providing them with insights into the efficacy of their teaching strategies and student learning patterns. Instructors have the ability to ascertain which educational materials, instructional strategies, or evaluation methods are most effective for their students. This enables educators to modify their instructional strategies in order to establish more productive and engaging learning environments.

Enhancing Student Retention Rates:

One critical aspect of EDM's responsibility to retain students in educational institutions is to identify the factors that contribute to student attrition or course abandonment. Educational institutions can employ predictive analytics to identify students who are at risk of academic challenges or disengagement and to provide targeted assistance, counseling, or mentorship as appropriate treatments. This increases the likelihood that students will complete their studies and enhances their level of satisfaction.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

Importance of Digital Learning Data:

Digital learning data has become more accessible as a result of the rapid advancement of online learning technology and Learning Management Systems (LMS). The abundance of educational data has resulted in the growing significance of educational data mining in contemporary educational institutions. EDM assists educational institutions in transforming unstructured student data into valuable insights that facilitate personalized learning, enhance lesson planning, and inform strategic decision-making process.

4. MACHINE LEARNING TECHNIQUES USED

a) Decision Tree

Many individuals utilize decision trees, a form of supervised machine learning, to predict student performance in online classes. It categorizes students into distinct groups according to their academic and behavioral characteristics, including quiz performance, attendance frequency, assignment completion, and level of participation. The method generates a hierarchical structure that is comprehensible and conducive to analysis. This renders it the optimal selection for evaluating schooling and forecasting achievement.

b) Random Forest

The Random Forest ensemble learning method enhances prediction accuracy and reliability by amalgamating many decision trees. It examines various aspects of a student, including their learning style, test performance, and preferred activities, to predict their academic success. Random Forest yields more precise and dependable predictions compared to individual decision tree models and mitigates the issue of overfitting.

c) Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a machine learning technique predominantly employed for categorizing tasks in educational data processing. Analyzing students' academic records and their online learning behaviors facilitates the categorization of learners into groups such as pass/fail, high-performing/low-performing, or at-risk/safe learners. Support Vector Machines (SVM) can precisely identify data and perform effectively with complex, high-dimensional educational datasets.

d) Naïve Bayes

Naïve Bayes is a probabilistic machine learning technique that employs statistical correlations between educational data and probability distributions to accurately forecast student performance in school. It examines factors such as quiz scores, assignment completion, attendance, and engagement in learning to assess a student's performance. The method is user-friendly, efficient, and proficient in handling extensive educational data within online learning platforms.

e) Artificial Neural Networks (ANN)

An artificial neural network (ANN) is a sophisticated machine learning model designed to emulate the architecture and operations of the human brain. Artificial Neural Network (ANN) models can identify intricate relationships among several facets of a student's learning, including academic achievement, engagement levels, learning preferences, and online activity patterns. Typically, these models are employed in intelligent learning systems to enhance predictive accuracy and analyze student performance.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

f) Deep Learning

Deep learning is a complex domain of machine learning that employs multiple layers of neural networks to analyze enormous volumes of intricate educational data. Deep learning methodologies autonomously identify concealed patterns inside online assessments, discussion forums, student activity logs, and instructional videos. These strategies enhance the accuracy of predictions in adaptive learning systems and facilitate personalized learning by assessing data and recommending intelligent actions.

5. RESULTS

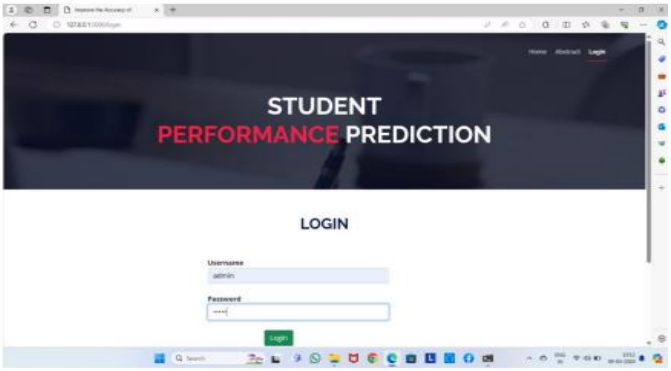


Fig.2. Login Page

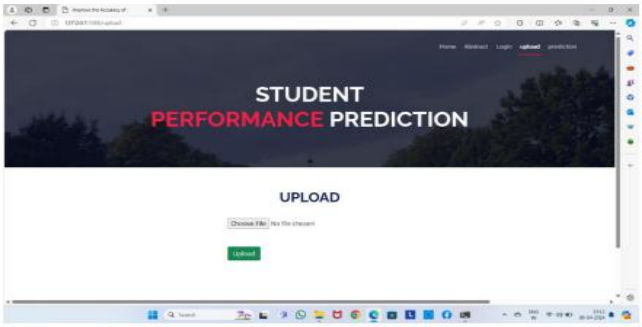
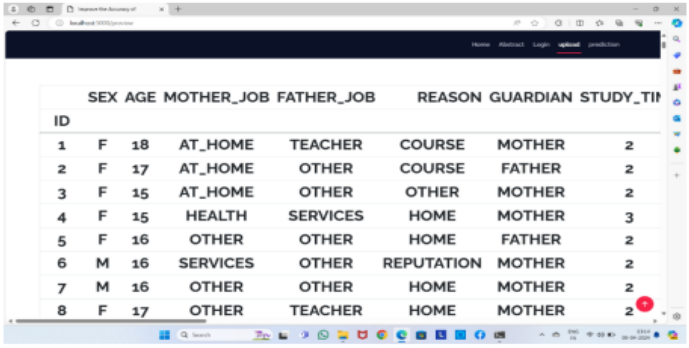


Fig.3. Upload Page



ID	SEX	AGE	MOTHER_JOB	FATHER_JOB	REASON	GUARDIAN	STUDY_TI
1	F	18	AT_HOME	TEACHER	COURSE	MOTHER	2
2	F	17	AT_HOME	OTHER	COURSE	FATHER	2
3	F	15	AT_HOME	OTHER	OTHER	MOTHER	2
4	F	15	HEALTH	SERVICES	HOME	MOTHER	3
5	F	16	OTHER	OTHER	HOME	FATHER	2
6	M	16	SERVICES	OTHER	REPUTATION	MOTHER	2
7	M	16	OTHER	OTHER	HOME	MOTHER	2
8	F	17	OTHER	TEACHER	HOME	MOTHER	2

Fig.4. Student Data View

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

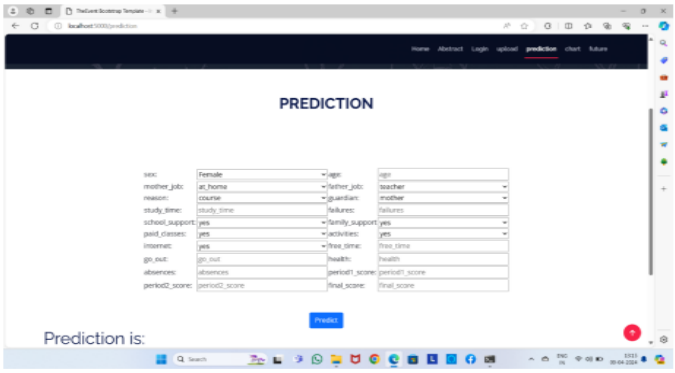


Fig.5. Prediction Page

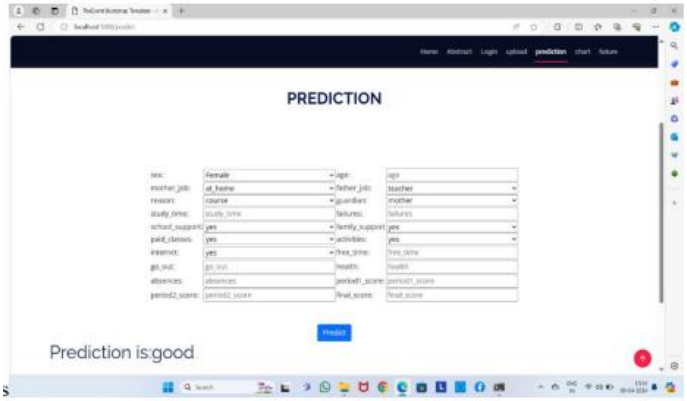


Fig.6. Prediction Result Page

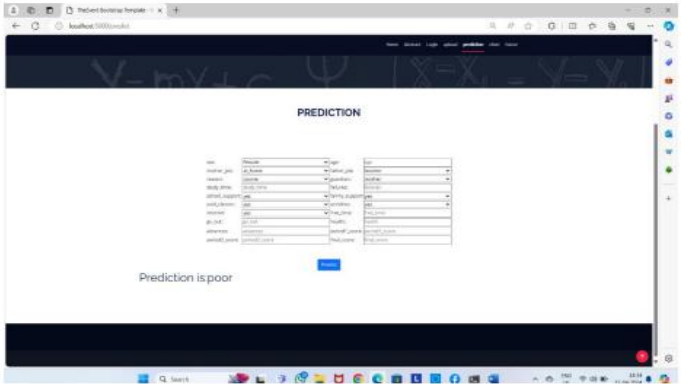


Fig.7. Poor Prediction Result

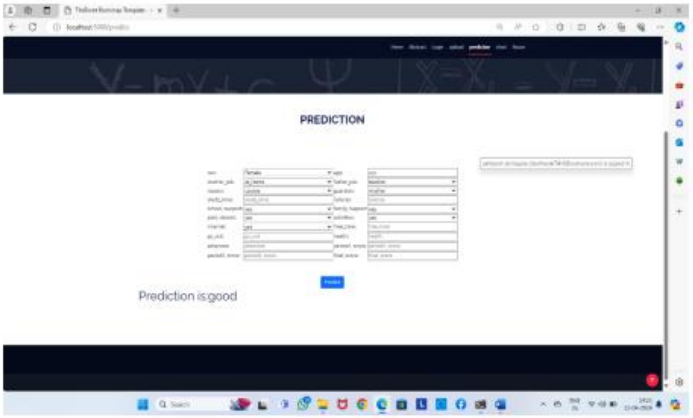


Fig.8. Good Prediction Result

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-008| Research Article
doi: https://doi.org/10.5281/zenodo.20675031

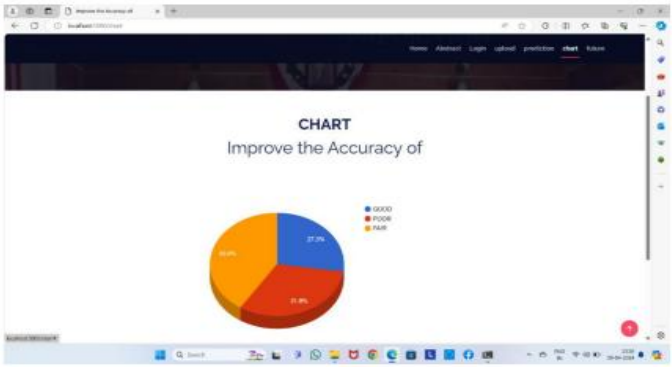


Fig.9. Accuracy Chart

6. CONCLUSION

In summary, the importance of data-driven intelligent technologies in improving student success and educational efficacy is demonstrated by research that utilizes machine learning to analyze and predict student performance in online courses. Schools can accurately predict academic success and identify at-risk students early by utilizing machine learning algorithms on student performance data, learning behaviors, attendance records, assessments, and engagement metrics. Consequently, educators and administrators can make more informed decisions, implement personalized learning strategies, and intervene at the most opportune times. The article explores the ways in which predictive analytics improves learning outcomes, reduces student attrition rates, and enhances the efficacy of flexible online education systems.Ultimately, the learning environment is improved by the application of machine learning in online education, which makes it more adaptable, student-centered, and effective. It also suggests the potential for future improvements in the surveillance of student achievement and the enhancement of technology in educational institutions.

REFERENCES

1. K. F. Hew and W. S. Cheung, “Students’ and instructors’ use of massive open online courses (MOOCs): Motivations and challenges,” Educ. Res. Rev., vol. 12, pp. 45–58, 2014.
2. M. Çetinkaya-Rundel, and D. A. Canelas, “Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers,” Comput. Educ., vol. 110, pp. 35–50, 2017.
3. J. Renz, F. Schwerer, and C. Meinel, “openSAP: Evaluating xMOOC Usage and Challenges for Scalable and Open Enterprise Education.,” Int. J. Adv. Corp. Learn., vol. 9, no. 2, pp. 34–39, 2016.
4. S. Li, Q. Tang, and Y. Zhang, “A Case Research on Learning Difficulties and Corresponding Supports for Learning in cMOOCs| Une étude de cas sur les difficultés d’apprentissage et le soutien correspondant pour l’apprentissage dans les cMOOC,” Can. J. Learn. Technol. Rev. Can. l’apprentissage la Technol., vol. 42, no. 2, 2016.
5. Namoun, Abdallah, and Abdullah Alshanqiti. "Predicting student performance using data mining and learning analytics techniques: a systematic literature review." Applied Sciences 11, no. 1 (2021): 237.

6. Saifuzzaman, M., Parvin, M., Jahan, I., Moon, N.N., Nur, F.N. and Shetu, S.F., 2021, June. Machine Learning Approach to Predict SGPA and CGPA. In 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST) (pp. 211-216). IEEE.
7. H. Alamri, Leena, Ranim S. Almuslim, Mona S. Alotibi, Dana K. Alkadi, Irfan Ullah Khan, and Nida Aslam. "Predicting Student Academic Performance using Support Vector Machine and Random Forest." In 2020 3rd International Conference on Education Technology Management, pp. 100-107. 2020.
8. Verma, C., Illés, Z. and Stoffová, V., 2019, February. Age group predictive models for the real time prediction of the university students using machine learning: Preliminary results. In 2019 IEEE International Conference on Electrical,ComputerandCommunication Technologies (ICECCT) (pp. 1-7). IEEE.
9. Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, “Mining student data using decision trees”, International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.