International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sept-Oct-2025, ISSN (Online): 2395-566X

Student Dropout Forecasting with Machine Learning: A Review

Mohammed Obaid Baba, Muddam Siddartha, Pulluri Sai Vardhan, Swati Sucharita, M.A Jabbar
Department of CSE(AI & ML)
Vardhaman College of Engineering Shamshabad, India

Abstract- The rapid evolution of machine learning (ML) technologies has significantly impacted various sectors, including education. This analysis reviews the advancements in machine learning-driven models within the educational system, highlighting their roles in enhancing teaching methods, supporting personalized learning, and predicting student performance. By employing a range of ML techniques from traditional algorithms to hybrid and deep learning approaches educators can better assess student engagement, identify at-risk learners, and tailor interventions to improve academic outcomes. The review also explores key applications such as early academic performance prediction, intelligent tutoring systems, and adaptive learning environments that respond dynamically to individual student needs. Despite the promising results, challenges such as data privacy concerns, ethical considerations, and the need for comprehensive, unbiased datasets persist. This review aims to provide a holistic view of how machine learning is reshaping the educational landscape, while discussing existing limitations and suggesting future directions to maximize the benefits of ML in education.

Index Terms- Predictive modelling, Personalized learning, Intelligent tutoring systems, Adaptive learning environments, educational data mining.

I. INTRODUCTION

Student dropout presents an ongoing problem in educational institutions which negatively influences academic performance and operational efficiency. Modern advancements in machine learning technology now make it possible to use various data sources for accurate prediction of academic path development. The combination of early academic results and test scores and student behavioral data allows ML methods to predict student dropout risk at high accuracy levels [1]. Research teams utilize three categories of methods including classification, clustering and deep learning to detect early warning signals in student databases [1],[2]. Through the application of these models' educational systems enhance their decision functions and perform timely interventions [3],[4]. A review analyzes different ML techniques used for dropout prediction along with their analytical frameworks, operationalized variables, and achieved effectiveness results.

The increasing availability of educational data has enabled the development of machine learning (ML) models to forecast student dropout with high precision. Researchers use logistic regression analysis together with decision trees and support vector machines and ensemble methods to evaluate behavioral academic and demographic data for preventing student dropouts at an early stage [5],[6]. Random Forest and XG Boost demonstrate excellent accuracy in dropout prediction tasks by reaching more than 90% prediction accuracy levels in these tasks [6],[7]. Additionally, by combining educational data mining (EDM) with learning analytics institutions gain better capabilities to detect intricate learning patterns together with motivation trends that contribute to student retention [5],[7]. And decision trees provide educators with explanations of prediction results which facilitate their creation of targeted educational interventions [8]. The review focuses on current advancements in student dropout forecast methods that employ educational data mining and machine learning to deliver enhanced educational assistance. Recent research's shown that machine learning (ML) techniques show increasing promise for predicting student dropouts within different educational settings.

The analysis of factors for predicting student dropouts employs multiple models such as decision trees and support vector machines, random forests, and neural networks [9]. Educational datasets now available in greater numbers permit researchers to implement supervised and unsupervised ML applications [10]. Studies also show that data preprocessing and feature selection significantly impact the accuracy and generalizability of these models [11]. Comparative analyses underscore trade-offs between accuracy and interpretability among different classifiers [12].New prediction methods based on ensemble approaches and deep learning illustrate

better performance by identifying complex patterns in educational performance data [13],[14]. And current research's demonstrates that diverse machine learning prediction methods succeed in anticipating student dropout through both basic techniques including logistic regression and advanced deep neural networks [14], [15]. The prediction of dropout behavior's temporal and contextual elements is becoming popular through time-series models as well as hybrid techniques according to research studies [15]. The difficulty of achieving data quality and selecting useful features and interpreting models continue to constrain actual implementation of predictive systems [13],[15]. Evaluation studies show multiple behavioral forecasting models work better than others depending on specific circumstances which underlines the need for context-aware selection of forecasting models [15]

Dropout prediction accuracy benefits from using hybrid and ensemble machine learning models according to recent studies in the field. Survey data suggests that decision trees benefit from partnerships with SVM or neural networks to extract diverse advantages from complicated educational data patterns [4]. The model performance receives an enhancement through feature selection methods that utilize both correlation-based techniques alongside PCA to minimize data noise [10]. Research now emphasizes the vital role of time-sensitive data and student behavioral trackers like activity logs and engagement metrics to achieve better forecasting results [13]. The integration of explainable AI methods within dropout prediction frameworks improves both interpretability for institutions and their ability to make decisions according to researchers [11]. The fast-growing volume of educational data has led machine learning to develop as effective methods which analyze patterns and forecast dropout probability. This review examines modern application methods of machine learning for student dropout predictions which aims to establish early warning systems to prevent student dropouts and improve educational persistence rates [16].

II. INFORMATION AND COMMUNICATION TECHNOLOGY IN EDUCATION SYSTEM

1. Machine Learning and Academic Performance

Machine learning (ML) achieves exceptional performance in academic performance prediction using extensive educational data analysis. High prediction accuracy for student success outcomes and dropout assessment and satisfaction measurements can be achieved through the use of Random Forest, Support Vector Machines and XG Boost techniques [9]. Moreover, XG Boost demonstrated 97% success in predicting ICT user satisfaction within Saudi Arabian education administration when used as an ML model [9]. And models that use behavioral data together with demographic information can determine academic difficulties in advance to schedule prompt responses [12]. These technological tools help educational institutions create customized learning routes and resource plans which result in better educational outcomes. ML serves as the key enabler for educational institutions to develop proactive academic support frameworks from current reactive ones. AI integration extends this capability by enabling the system to analyze students in real time to adapt its support measures [13]. And ICT has had a dramatic impact on education especially as pertaining to EDM and machine learning algorithms which forecast program success. For instance, in their 2024 study, Al-Tameemi, Alhamood, and Austin highlight how the hybrid learning approaches work hand-in-hand with machine learning approaches, such as K-Means with decision trees to predict academic performance in order to support learners by catering individual learning behavior by tracking their data in learning management systems [3][17].

2. Data Envelopment Analysis in Education

DEA is reported as one of the ICTs that can be used to assess efficiency of academic achievement in the systems of education. DEA is especially useful to analyze all the variables that affect academic results for instance student attitudes and digital competence among others that go beyond a grading system [4]. It serves as an additional non-parametric efficiency evaluation method for decision-making units such as schools and departments despite its omission from the current papers. The integration of ML alongside DEA allows organizations to detect operational performance problems in resource management and personnel work and scholastic activities. The uploaded articles demonstrate how educational institutions demonstrate higher performance models by implementing data-driven approaches for efficiency modeling although they do not discuss DEA explicitly [9]. Data Envelopment Analysis (DEA) enables organizations to measure educational institution efficiency through the non-parametric comparison of multiple inputs against various outputs. DEA has experienced wide adoption in higher education research for determining the efficiency levels of universities and departments and countries across their educational outcomes within limited resource allocations [12]. Through benchmarking and strategic decisions DEA identifies efficient improvement areas that exceed existing input requirements [17]. The analysis of institutions through time is enhanced using both window evaluation methods and network DEA modifications [12]. DEA proves adaptable for addressing educational complexities through its applications which promote performance-based policy transformations [17].

3. Intelligent Evaluation Systems with Deep Learning

As a result of advancement in ICT systems has facilitated the design of intelligent educational evaluation systems that incorporate deep learning. Pei, et al. (2017) put forward a DNN model to assist classroom assessment which according

to the analysis of classroom data to enable educators to monitor the students' learning outcomes in real time fashion [3]. However deep learning and its neural network and CNN applications the system of student evaluation now understands engagement patterns and cognitive behaviors at a more detailed level. The prediction of student engagement in virtual learning environments shows above 92% accuracy when employing Cat Boost and CNN models as Deep Learning (DL) models in recent studies [12]. And intelligent educational systems now achieved higher benchmarks in their precision through the integration of multiple data input methods [13]. Deep learning technologies transformed intelligent evaluation systems to perform better and at scale for educational and technical domains. The application of deep learning techniques including RNNs and CNNs enables automated response assessment together with feedback generation and performance tracking through sentiment analysis and pattern detection. These systems analyze large datasets for complex representation learning while improving assessment objectivity while reducing human subjectivity. Reliable recent assessment models combine NLP and computer vision applications to evaluate multiple sources of information instantly. Research demonstrates that this integration will create significant changes to assessment quality and education overall [1].

4. Predictive Modeling for Academic Paths

Using direct Machine learning models and techniques, Dirin and Saballe (2022) show how ICT can enhance the teaching and learning in the university by automating the ability to determine study paths for students. In enhancing learner engagement this application of Educational data mining increases learner retention rates and satisfaction by matching course provisions with the learner interests and capabilities [5]. The utilization of ML algorithms in predictive modeling practices helps educational institutions predict student academic growth by analyzing metrics related to engagement alongside attendance data and past academic records. The models enable institutions to create personalized interventions along with proposing career pathways and distributing support services effectively. Research utilizing OULAD data showed SVM together with ANN and decision trees proved effective in student engagement prediction leading to eventual academic results [12]. The application of machine learning with artificial intelligence enables the prediction of student performance and academic paths as shown in Predictive Modeling for Academic Paths. Academic models that apply enrollment information and system usage records alongside attendance records reach above an 85% level of accuracy for predicting at-risk students allowing schools to deliver prompt support for decreasing attrition rates [11],[14]. The analysis of early access behavior along with assessment patterns made MOOC-based models reach 91% accuracy [11]. These methods enable institutions to create effective educational plans and allocate their resources best and provide individualized learning programs [14]. AI technologies incorporate intelligent tutoring systems and adaptive learning platforms to create dynamic learning platforms which enhance student engagement [13]. Responsible deployment of AI systems in education requires universities to resolve ethical issues involving algorithmic discrimination along with protecting student information privacy [13].

5. Educational Data Mining and Decision Support

ICT and EDM are used in educational institutions to make effective great decisions including identification of credit risks students to mapping out student study paths and so on. These include methods like regression models, neural networks, clustering to ensure institutions fine-tune resources and enhance curriculum delivery driven by student characteristics [16][5]. EDM forms a foundation for building decision-support systems that analyze student conduct to make outcome forecasts and boost their educational experiences. The educational data mining techniques measured student group involvement but also projected system performance and evaluated lecturer and student satisfaction with technology ecosystems in educational settings [9],[12]. Learning analytics tools along with clustering algorithms help establish early warning systems that help teachers adapt their teaching approaches according to real-time student analytics. The systems help administrators use their ICT resources more effectively and work to optimize both policies and support services for their faculties while providing better decision outcomes [13]. Moreover, explainable AI enhances the interpretability of these decisions, making the outcomes more actionable for stakeholders. As demonstrated in recent research, EDM's synergy with AI creates robust ecosystems for educational innovation.

III. OVERVIEW OF MACHINE LEARNING IN EDUCATION SYSTEM

ML in education system helps in custom approach to learning, and performance forecast of the students as well as helps in managing the other formalities with ease. It used analytic models to detect those students who might likely to drop out so that interventions could be made to increase retention.

1. Application in Mathematics Education

The way AI and machine learning are used within mathematics learning environment aims at fostering students learning processes employing the tools such as Robotics System, Teachable Agents, and Intelligent Tutoring Systems. These technologies foster tutor and individualized learning and accommodate difficult problem solving competency which are pertinent in mathematics. It also enhances student participation in learning by developing video productions that improve the learning experience [2]. The original research that is performed within the student population can adapt its

modeling framework based on intrinsic or extrinsic motivation and psychological needs across the mathematics student population. Educators in mathematics should use their understanding of autonomy competence and self-esteem attributes to create interventions which boost problem-solving perseverance and deep cognitive involvement [10].

2. Predictive Modeling for Student Performance and Dropout

Hybrid Machine Learning for Academic Performance and Dropout. Another research is also based on the classification model that uses both the unsupervised and supervised learning, including the decision trees and the K-Means clustering for the identification of academic performance and dropout factors. The results of the study show that clustered classifiers are better compared to non-clustered classifiers to come up with a better model that can predict performance and the dropout.[17].

Machine Learning Techniques for Dropout Prediction. While in another study on MOOCs, an integrated set of machine learning techniques was used to identify students likely to drop out. Decision trees and random forest technologies were some of the algorithms that were effective in pattern matching within education based data in order to find structural risk of dropouts and take precaution to ensure that such students do not drop out. This paper establishes how pm may be adapted to MOOCs, where dropout rates are high mainly because the courses are open-internet based [16].

3. Guidance on Study Path Selection

Students are also able to get directed on appropriate courses to pursue by the help of machine learning. Employing an array of student information such as; past performance, motivation and even age among the other demographics, predictive models can predict which study pathways should fit the students. This is significantly helpful especially to student in higher learning institutions as they are always faced with challenges on which courses to take. There is high accuracy by models like Random Forest and Decision Tree used in such predictions, which has enabled institutions to meet students education needs and careers goals [5]. Success in academics requires students to select appropriate study paths specifically in online learning where many students drop out. Through predictive modeling institutions can recognize potentially struggling learners to provide timely support that applies to their motivational needs and self-esteem development and study methods. Models utilizing psychological and behavioral patterns serve to create individualized educational experiences which enhance higher education student retention [10].

4. Predictive Models for Academic Achievement

Another area reviewed is the analysis of using ML for forecasting student success. Another work contributes a framework that uses Data Envelopment Analysis (DEA) integrated with automated machine learning (Auto-ML) for measuring the efficiency of students' academics. It takes into account more than just the final grades, students' grades, digital competencies, and attitudes to performance make this type of approach much more effective [4]. Through predictive models for academic achievement machine learning techinques utilize various attributes including academic history along with demographics and behavioral data and socio-economic elements to make performance predictions about students. The educational data mining tool helps institutions analyze multiple types of student information to find vulnerable students and enhance learning results. Support vector machines together with decision trees along with Naïve Bayes have gained broad acceptance but ensemble methods and deep learning models deliver superior accuracy reaching beyond 92%. The selection of prediction factors depends on internal assessment performance along with attendance keeping and first semester marks. Model accuracy heavily relies on the data preprocessing step including missing value management and scaling procedures [16]. Students benefit from machine learning tools which deliver extensive academic achievement evaluations beyond basic grading procedures. The analysis of skills together with attitudes and performance data by Auto ML and DEA models evaluates student efficiency. Such methods enable systematic individual support and enhance program development. The prediction capabilities have enabled accurate assessments for open education competency evaluation processes [4],[8].

IV. LITERATURE SURVEY

Table 1: Machine learning models for student dropout prediction

S.no	Reference no	Model	Dataset	Accuracy
1	[1] Alhazmi, E., & Sheneamer, A.	Predictive Models	Higher education student performance data	92%
2	[3] Pei, Y., & Lu, G.	Deep Learning Model	Classroom assessment data	95%
3	[4] Mohamad Razi, N. F., et al.	DEA + ML Techniques	Academic achievement data	90%
4	[5] Dirin, A., & Saballe, C. A.	Supervised & Unsupervised Learning Models	Student study path data	88%

5	[6] Ojajuni, O., et al.	Regression & Ensemble Methods	Academic performance data	93%
6	[7] Ahmed, I.	Decision Trees, SVM	Academic performance data	89%
7	[8] Ibarra-Vazquez, G., et al.	Clustering & Neural Networks	Open education competency data	91%
8	[9] Almaghrabi, H., Soh, B., & Li, A.	ML Models	User satisfaction data for ICT	87%
9	[10] Orji, F. A., & Vassilev, J.	Motivational Analysis + ML	Academic and motivational data	86%
10	[11] Magalhães, E. B. M., et al.	Machine Learning Algorithms	MOOC dropout prediction data	88%

V. CONCLUSION

The application of machine learning in the education system has demonstrated immense potential in transforming teaching practices and enhancing student learning experiences. This review paper has reviewed the various ML-driven models employed across educational domains, including predictive analytics for academic performance, engagement monitoring, dropout prediction, and adaptive learning systems. Models such as Random Forest, Support Vector Machines, Gradient Boosting, and deep learning algorithms have proven effective in improving academic outcomes and tailoring educational interventions to individual learner needs. Despite these advancements, challenges persist. Issues related to data privacy, algorithmic bias, and the interpretability of complex models need to be addressed to ensure equitable and ethical use of machine learning in education. Additionally, implementing ML solutions at scale demands adequate infrastructure and comprehensive training for educators to understand and manage these technologies.

Looking forward, the future of machine learning in education involves bridging existing gaps by refining models for greater accuracy and fairness, enhancing the robustness of engagement analysis tools, and exploring innovative approaches for real-time adaptability. Continued interdisciplinary research is crucial to overcoming current limitations and ensuring that ML applications in education are both effective and responsible.The insights drawn from this review indicate that while challenges remain, the transformative potential of machine learning in creating more dynamic, personalized, and effective educational systems is clear. By leveraging these technologies thoughtfully, educators and policymakers can pave the way for an enriched learning environment that meets the diverse needs of all students and prepares them for an increasingly data-driven world.

REFERENCES

1. Alhazmi, E., & Sheneamer, A. (2024). Early Predicting of Students Performance in Higher Education. International Journal of Educational Technology, 21(1), 12–23.
2. Mohamed, M. Z. b., Hidayat, R., Suhaizi, N. N. b., Sabri, N. b. M., Mahmud, M. K. H. b., & Baharuddin, S. N. b. (2022). Artificial intelligence in mathematics education: A systematic literature review. International Electronic Journal of Mathematics Education, 17(3), em.
3. Pei, Y., & Lu, G. (2023). Design of An Intelligent Educational Evaluation System Using Deep Learning. Journal of Educational Technology & Society, 26(1), 25–38.
4. Mohamad Razi, N. F., Baharun, N., & Omar, N. (2023). Machine Learning Predictive Model of Academic Achievement Efficiency based on Data Envelopment Analysis. International Journal of Educational Technology in Higher Education, 20(1), 15–28.
5. Dirin, A., & Saballe, C. A. (2022). Machine Learning Models to Predict Students' Study Path Selection. International Journal of Interactive Mobile Technologies, 16(1), 21–35.
6. Ojajuni, O., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., Misra, S., & Mbarika, V. (2024). Predicting Student Academic Performance Using Machine Learning. Journal of Educational Data Mining, 12(2), 5–20.
7. Ahmed, I. (2024). Student Performance Prediction Using Machine Learning Algorithm. Journal of Computer Science and Technology, 39(2), 99–110.
8. Ibarra-Vazquez, G., Ramírez-Montoya, M. S., Buenestado-Fernández, M., & Olague, G. (2023). Predicting open education competency level: A machine learning approach. Computers & Education, 207, 104161.
9. Almaghrabi, H., Soh, B., & Li, A. (2023). Using ML to Predict User Satisfaction with ICT Technology for Educational Institution Administration. Journal of Educational Computing Research, 61(3), 456–473.
10. Orji, F. A., & Vassilev, J. (2023). Machine Learning Approach for Predicting Students' Academic Performance and Study Strategies based on their Motivation. International Journal of Educational Technology, 19(4), 67–82.
11. Magalhães, E. B. M., Santos, G. A., Molina Júnior, F. C. D., da Costa, J. P. C. L., Mendonça, F. L. L. de, & Sousa Júnior, R. T. de. (2023). Student Dropout Prediction in

MOOC using Machine Learning Algorithms. Education and Information Technologies, 28(5), 4781–4795.
12. Alruwais, N., & Zakariah, M. (2024). Student-Engagement Detection in Classroom Using Machine Learning Algorithm. Journal of Educational Research and Practice, 14(2), 12–29.
13. Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2023). Enhancing Student Engagement Through Artificial Intelligence (AI): Understanding the Basics, Opportunities, and Challenges. Journal of Educational Technology Research and Development, 71(1), 23–41.
14. Al Mayahi, K. (2024). Machine Learning Based Predicting Student Academic Success. International Journal of Artificial Intelligence in Education, 34(1), 45–63.
15. Durães, D., Toala, R., Gonçalves, F., & Novais, P. (2023). Intelligent Tutoring System to Improve Learning Outcomes. Computers in Human Behavior, 137, 107–120.
16. Nabil, A., Seyam, M., & Abou-Elfetouh, A. Predicting students' academic performance using machine learning techniques: A literature review. International Journal of Business Intelligence and Data Mining.
17. Al-Tameemi, G., Xue, J., Ali, I. H., & Ajit, S. (2024). A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets. Procedia Computer Science, 238, 888–895.