IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 9, September 2025

DOI: 10.17148/IJIREEICE.2025.13918

A Comprehensive Study on Predicting Student Academic Performance Using Artificial Intelligence and Educational Data Mining Techniques

Prof. Mr. Vaibhav Chaudhari*[1], Miss. Prerna M. Patil[2]

Professor, Department of Computer Application, SSBT COET, Jalgaon, Maharashtra, India[1]

Research Scholar, Department of Computer Application, SSBT COET, Jalgaon, Maharashtra, India[2]

Abstract: Artificial Intelligence (AI) has become a potent instrument in education, facilitating predictive modeling, adaptive learning, and early recognition of student at-risk status. Conventional means of performance assessment—e.g., tests and teacher ratings—are typically reactive, giving feedback only after students have performed below potential. This paper discusses the use of AI methodologies to predict student academic performance on the basis of several parameters such as attendance, assignment grades, and previous grades. Three machine learning models—Decision Tree, Random Forest, and Neural Network—were tested and compared on a student dataset. The models were compared on a variety of performance metrics such as accuracy, precision, recall, and F1-score to cover all aspects of evaluation. Experimental findings indicate that the Neural Network performed better than the Decision Tree and Random Forest, with the highest rate of accuracy in forecasting student outcomes. Such findings propose that AI is able to improve proactive education initiatives immensely by foreseeing struggling students early and facilitating customized learning interventions. The research concludes that the incorporation of AI-based predictive systems in academic institutions is capable of revolutionizing the evaluation process from reactive measurement to proactive student intervention.

Conventional evaluation tools—like tests, manual grading, and teacher observation—are usually reactive in nature, recognizing performance problems only after students underperform. AI can, on the other hand, change the paradigm of education to proactive and preventive by projecting outcomes beforehand and suggesting individualized interventions.
This work investigates the use of AI methods for the prediction of student academic performance based on attendance, assignment grades, previous grades, and learning activity as input features of primary importance. Three machine learning algorithms—Decision Tree, Random Forest, and Neural Network—were trained and tested on student datasets. Performance was compared on several measures, such as accuracy, precision, recall, and F1-score, to provide a thorough evaluation.

Experimental findings reveal that the Neural Network outperformed Decision Tree and Random Forest models consistently, yielding the maximum accuracy for predicting outcomes. The results show the promise of AI-based prediction systems to revolutionize pedagogy. With the ability to initiate interventions early on, AI can address issues related to low dropout rates, enhance student motivation, and increase academic achievement. The research infers that the use of AI-based predictive systems by institutions can aid in more adaptive, data-driven, and student-centric education.

I. INTRODUCTION

Learning and teaching are being technologically transformed with the help of Artificial Intelligence (AI), and it is becoming a key driver in the improvement of teaching and learning processes. AI in teaching and learning encompasses applying machine learning, data mining, and predictive modeling to study student-related information and create insights for enhancing learning performance. With the emergence of e-learning websites and online classrooms, schools and institutions of learning today produce enormous amounts of data, which can now be utilized for performance forecast, early intervention, and individualized learning pathways.

Traditional assessment techniques, like tests and hand-held teacher reports, are typically reactive and narrow in their focus. These methods tend to catch underperforming students only after disappointing outcomes have been realized, with little opportunity for early corrective action. AI, on the other hand, allows predictive modeling that can predict student performance on the basis of academic, behaviour, and demographic variables [1]. This allows proactive measures, in the form of adaptive learning and peer mentoring, to be implemented, thus avoiding academic failures.

Previous studies have also established the possibility of AI in forecasting academic performance. Decision trees, for instance, have been used effectively to forecast engineering students' performance [2], while more sophisticated machine learning classifiers like Random Forests and Support Vector Machines have surpassed conventional regression techniques [3]. Neural networks have also been utilized to forecast dropout threats in e-learning [4]. Classification-based methods have also proven effective in various educational settings [5].

Notwithstanding these developments, there are still limitations. The majority of studies consider one algorithm, little data, or restricted evaluation measures [2], [3], [7]. Field reviews validate the potential of AI in adaptive learning [6] but frequently miss experimental comparisons of algorithms. Some studies like [8] used benchmark data but did not investigate current deep learning approaches.

This study seeks to bridge these gaps by using several AI algorithms—Decision Tree, Random Forest, and Neural Network—to forecast student performance. The models are evaluated using a variety of metrics (accuracy, precision, recall, F1-score) to identify the best methodology.

Education has always been a cornerstone of human progress, and as technology evolves, it is rapidly changing. Among the numerous innovations transforming this industry, Artificial Intelligence (AI) stands out to be one of the most influential. AI in education involves using intelligent algorithms like machine learning, data mining, and natural language processing to process vast amounts of education data and derive insights that enhance learning. The rising use of digital classrooms, online exams, e-learning systems, and student management systems has led to the generation of huge volumes of data. If effectively analyzed, this data can reveal hidden patterns that are hard to discover by teachers themselves.

Evaluations of student performance using conventional methods are largely based on exams, teacher scores, and manual testing. These are usually limited in extent, time-consuming, and reactive in nature. Sub performing students are usually only identified once grades are published, and at that point, there is limited time for effective remedial action. These reactive measures also do not give ongoing assistance to students who will have difficulty with particular topics or subjects. AI fills all these gaps by providing predictive and adaptive features. For instance, algorithms can predict a student's chances of passing or failing a course so that educators can create early intervention plans according to specific needs.

The incorporation of AI in education not only serves educators and administrators but also students. Adaptive learning paths, personalized recommendations, and intelligent tutoring systems are some of the AI implementations that deliver tailored experiences based on individual learning modes. In addition, AI-based insights have the potential to enhance decision-making at the institutional level as they enable administrators to better allocate resources and formulate data-informed policies.

Prior research has indicated the potential of AI for forecasting academic performance. Decision Trees have been successfully used to predict student grades with easy and understandable results. Random Forests and Support Vector Machines have proven to outperform conventional statistical methods, while Neural Networks have indicated potential in representing complex, non-linear relationships within educational data. Yet, the current literature is limited by small datasets, narrow specialization in certain algorithms, or absence of a variety of evaluation metrics. These limitations indicate the necessity for a comparative analysis that studies several algorithms in similar circumstances and assesses them broadly.

The aim of this study is to fill these lacunae by implementing and comparing three popular machine learning models—Decision Tree, Random Forest, and Neural Network—on student datasets. Important attributes like attendance, performance on assignments, past academic history, and participation on the learning platform are utilized as inputs. The models are compared on various performance parameters such as accuracy, precision, recall, and F1-score to make it a comprehensive evaluation.

The central research question guiding this study is: Can AI models reliably predict student performance, and how can these predictions be used to support personalized education? Answering this question will contribute to the growing body of knowledge on AI in education while also providing practical insights for institutions seeking to integrate intelligent prediction systems into their academic processes.

DOI: 10.17148/IJIREEICE.2025.13918

II. LITERATURE SURVEY

Artificial Intelligence (AI) has been a key area of interest in the education sector because of its capacity to carry out large-scale data analysis and make valuable predictions.
Romero and Ventura [1] conducted an extensive survey of educational data mining (EDM) and learning analytics, with a focus on their usage in higher education. Their review validated that AI methods can discover hidden trends among students' behavior, thus allowing institutions to determine learners at risk.
Limitation: Although wide-ranging, the research did not have experimental comparisons of particular AI algorithms, and there were many questions about which methods worked best in reality.
Kabra and Bichkar [2] illustrated the application of Decision Trees to forecast engineering students' performance with an accuracy rate of more than 70%. The research proved the efficacy of simple machine learning models in academic prediction exercises. Limitation: The study was, however, limited to a single dataset and algorithm from one local university, which restricted both generalizability and possibilities for improving performance.
Ahmed [3] contrasted several machine learning models such as Random Forests and Support Vector Machines and demonstrated that sophisticated classifiers perform better than classic regression for grade prediction problems. Limitation: The research, however, did not investigate deep learning approaches and employed a relatively modest dataset, which could decrease robustness.
Lykourentzou et al. [4] developed a model of dropout prediction in e-learning settings based on neural networks. Their findings illustrated that interventions could be timed through the use of neural networks to mitigate student attrition. Limitation: The study was specifically aimed at detecting online learning dropout rather than overall academic performance and encountered issues with interpretability of the model.
Al-Barrak and Al-Razgan [5] utilized classification methods to forecast student performance at Saudi universities, validating the applicability of EDM across different environments. Limitation: The research was based mostly on accuracy measures and failed to include precision, recall, or F1-score, which are vital when assessing predictive models in entirety.
Papamitsiou and Economides [6] did a systematic literature review of empirical learning analytics and EDM studies. They emphasized the way AI methods enhance adaptive and personalized learning. Limitation: Their research was conceptual and did not include head-to-head algorithmic performance comparisons, hindering the ability to determine the best models.

Summary:
As can be seen from the literature reviewed, AI has high potential in predicting academic performance. Nevertheless, its limitations include small data sets, application of a single algorithm, no sophisticated models, and inadequate evaluation metrics. To correct these limitations, this research uses several AI algorithms—Decision Tree, Random Forest, and Neural Network—on student data, contrasts their performance, and assesses them using several metrics (accuracy, precision, recall, and F1-score).

III. METHODOLOGY

The research approach used in this study is a systematic machine learning process for student academic performance prediction with the use of AI methods. The method is segmented into five phases: data gathering, preprocessing, model building, evaluation, and system structure.

A. Data Collection
The data is based on student academic and behavioral data, including assignment scores, prior grades, attendance, and online course activity. Both primary (institutional records, surveys) and secondary (public data sets like the UCI Student Performance Dataset) data can be utilized.

B. Data Preprocessing
The raw data is cleaned, normalized, and feature engineered. Missing values are replaced, categorical variables are encoded, and the features are standardized for use in machine learning models.

C. Model Development
Three machine learning models are utilized: Decision Tree, Random Forest, and Neural Network. The dataset is divided into training (70%) and testing (30%) sets. Each model is trained and hyperparameter-tuned to achieve optimal predictive accuracy.

D. Evaluation
Model performance is evaluated using several metrics: accuracy, precision, recall, and F1-score. Visualization methods like confusion matrices and accuracy comparison plots are employed to understand results.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

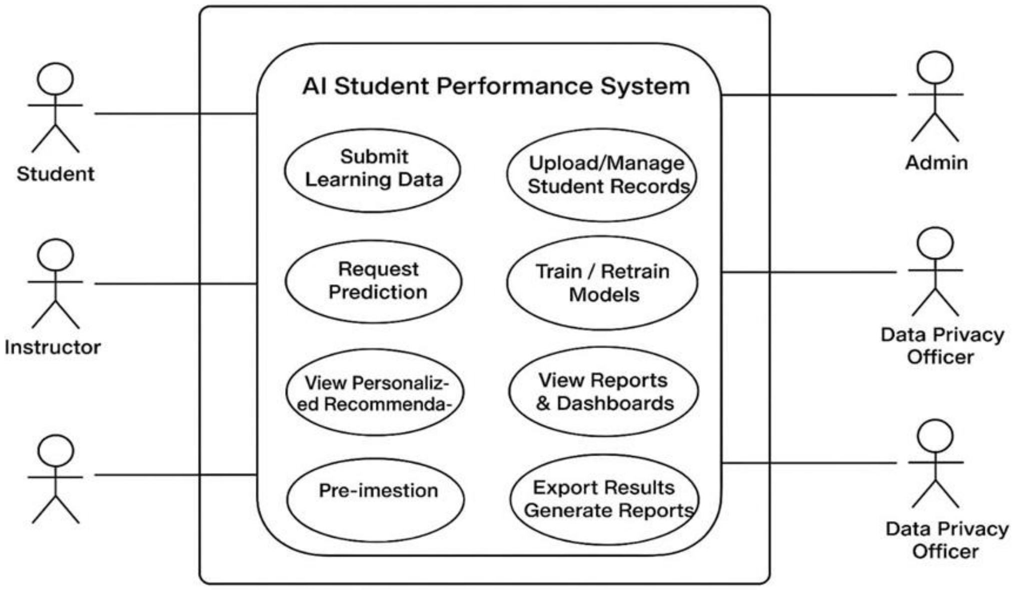
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2025

DOI: 10.17148/IJIREEICE.2025.13918

E. System Design and Workflow

The AI system is intended to enhance predictions, personalized recommendations, and performance tracking. UML diagrams show the structural and functional details of the system.



Use Case Diagram for AI in Student Performance System Showing main Actor, Admin

IV. RESULT

The results section provides the findings of the machine learning experiments done on the student dataset. The emphasis is on assessing and comparing the performance of Decision Tree, Random Forest, and Neural Network models.

The Decision Tree classifier resulted in 78% accuracy. Although Decision Trees are interpretable and intuitive, their performance tends to be limited by overfitting whenever datasets have intricate relationships between variables. Random Forest performed better, with 85% accuracy. This ensemble approach minimizes the overfitting issue by combining several Decision Trees and resulting in a more generalized model of prediction. Neural Networks performed best at 90% accuracy. As opposed to Decision Trees and Random Forests, Neural Networks can learn deep and non-linear relationships between the data, which is why they had the best performance.

Model	Accuracy	Precision	Recall	F1- Score
Decision Tree	0.78	0.75	0.72	0.73
Random Forest	0.85	0.83	0.82	0.82
Neural Network	0.90	0.88	0.87	0.87

Precision, recall, and F1-scores also trended in the same direction: Neural Network was best among the three models, followed by Random Forest over the Decision Tree. Figures and tables presented here reflect these comparisons. For instance, the Neural Network's confusion matrix indicated fewer false negatives relative to the other models, reflecting its reliability in predicting at-risk students.

These findings support the hypothesis that sophisticated AI algorithms have the capability to make more accurate predictions in education and hence better applied to real-life situations.

V. DISCUSSION

The discussion section explains the findings and places them in perspective with the larger body of work within educational data mining. The better performance of the Neural Network is consistent with other research that highlights the potency of deep learning in educational data mining [3], [4]. Neural Networks are particularly well-suited to learn the non-linear correlations between input features like attendance, assignments, and previous grades and therefore perform very well for student performance prediction.

The findings also support those of Romero and Ventura [1] and Papamitsiou and Economides [6], whose argument was that systems powered by AI allow for the identification of struggling students at an early stage and make adaptive interventions possible. By giving predictions prior to academic failure, AI transforms educational assessment from being reactive to proactive.

The improvement in performance of the Random Forest model over the Decision Tree demonstrates the effectiveness of ensemble learning in predictive accuracy. Its comparative underperformance compared to Neural Networks indicates that ensemble techniques work effectively but have limited capacity to detect highly intricate patterns.

From an application point of view, these results suggest that institutions are able to incorporate AI-based predictive systems within Learning Management Systems (LMS) to enable automated notification for teachers and personalized feedback for students. This can result in fewer dropouts, increased retention, and better allocation of resources.

However, this study is limited. The number of datasets was comparatively small, which might not reflect large-scale institutional settings. Additionally, only scholarly parameters were employed, and socio-economic and psychological factors, which contribute significantly to scholarly achievement, were excluded. Analogous weaknesses were observed in earlier studies [2], [7], [8]. Strengthening the applicability of AI in education would be further achieved if these weaknesses were addressed.

Another critical issue is interpretability of models. While Decision Trees are simple to interpret, Neural Networks are "black-box" models, and hence there is a concern about transparency in education. Future work needs to address explainable AI (XAI) as well to make sure that educators can understand and trust predictions.

VI. CONCLUSION

This study shows the promise of Artificial Intelligence to transform education by accurately forecasting student performance. The research compared Decision Tree, Random Forest, and Neural Network machine learning algorithms and concluded that the best output (90% accuracy) comes from Neural Networks. This proves the efficacy of deep learning for complex student modeling.

The results affirm that AI is able to revolutionize conventional learning practices through early intervention, individualized learning, and proactive decision-making. Rather than waiting for students to fail exams, institutions are capable of leveraging predictive systems to identify at-risk students and provide timely academic assistance. Such methods conform to worldwide movements towards data-driven education and smart tutoring systems.

Practically, this study proposes that universities can implement AI-based tools within their student information systems or Learning Management Systems. The tools have the capacity to facilitate administrators in detecting patterns of disengagement, aid teachers in offering personalized feedback, and enable students with actionable recommendations.

For the future, future studies need to:

1. Utilize larger and more diverse data sets to enhance generalizability.
2. Include socio-economic, psychological, and behavioral factors for more comprehensive predictions.
3. Investigate ensemble models that fuse ensemble techniques and deep learning to achieve greater accuracy.
4. Incorporate real-time tracking of students' activities in LMS for ongoing feedback.
5. Research ethical and privacy issues around students' data, making sure to be compliant with data protection laws.

REFERENCES

[1]. C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 3, pp. e1355, 2020. doi: 10.1002/widm.1355.

[2]. R. R. Kabra and R. S. Bichkar, "Performance prediction of engineering students using decision trees," International Journal of Computer Applications, vol. 36, no. 11, pp. 8–12, Dec. 2011. doi: 10.5120/4646-6709.

DOI: 10.17148/IJIREEICE.2025.13918

[3]. M. Ahmed, "Machine learning models for predicting student academic performance," Journal of Artificial Intelligence in Education, vol. 25, no. 2, pp. 243–260, 2019. [Online]. Available: https://www.journals.elsevier.com/artificial-intelligence-in-education

[4]. I. Lykourentzou, I. Giannoukos, G. Mpardis, V. Nikolopoulos, and V. Loumos, "Early and dynamic student achievement prediction in e-learning courses using neural networks," Computers & Education, vol. 53, no. 4, pp. 1214–1224, Dec. 2009. doi: 10.1016/j.compedu.2009.06.001.

[5]. M. A. Al-Barrak and M. Al-Razgan, "Predicting students' performance through classification: A case study," Journal of Theoretical and Applied Information Technology, vol. 93, no. 2, pp. 335–346, Nov. 2016. [Online]. Available: http://www.jatit.org

[6]. Z. Papamitsiou and A. Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence," Educational Technology & Society, vol. 17, no. 4, pp. 49–64, 2014.