



Student Performance Prediction in Online Courses Using Machine Learning Algorithms

S. Akshitha¹, T. Abhinaya², U. Sreekanth³, S. Siva Shankar⁴

^{1,2,3,4}Department of CSE (Data Science), Tadipatri Engineering College, Tadipatri.

Abstract:

The rapid expansion of online education platforms has generated vast amounts of learner data, providing new opportunities to predict and enhance student performance through data-driven methods. Using the Random Forest algorithm, this study suggests a machine learning-based method for forecasting online course success. To determine the main elements affecting academic results, a variety of features are examined, including learning activity logs, assessment results, attendance trends, and engagement indicators. Because of its high accuracy, robustness, and capacity to manage intricate, nonlinear interactions within the dataset, the Random Forest model is used. According to experimental results, the model successfully divides students into performance groups, allowing for the early detection of learners who may be at danger. The results demonstrate how Random Forest-based prediction algorithms can help teachers and educational institutions create individualized learning plans, increase student retention, and boost academic achievement in general in online learning settings.

Keywords: Prediction, Machine learning algorithms, courses, academic performance, students.

I. INTRODUCTION

Online learning has become very popular, but many students drop out before finishing their courses. This is a big problem because it affects students' goals and wastes resources. To help solve this, we can use computer programs that learn from data (called machine learning) to predict which students might struggle or drop out. These predictions can help teachers support students early on. In this study, we use machine learning to predict student performance in online courses by looking at how students engage with the course, their background, and past grades. We'll compare different machine learning methods to see which one works best, and identify what factors are most important for predicting student success. This can help make online learning more effective and improve outcomes for students. To make accurate predictions, we'll look at various things like how often students log in, watch videos, do assignments, and their scores on quizzes. We'll also consider student demographics like age and location, and their past academic performance. By analyzing all this data with machine learning, we aim to create a tool that helps teachers identify students who need extra help. This way, teachers can provide targeted support, like extra tutoring or personalized feedback, to improve student outcomes and reduce dropout rates. If we can predict which students might struggle, teachers can help them right away. This can help more students pass and finish their courses. We're using machine learning to look at student data and find patterns that can help teachers support students better. We'll use machine learning algorithms like Random Forest and Logistic Regression to predict student performance. These algorithms are good at finding patterns in student data and telling us which students might need help. By using these predictions, teachers can support students early and improve outcomes.

Online courses have become very popular, but many students face difficulties such as lack of motivation, poor time management, or limited interaction with instructors. Because of this, some students fall behind or drop out before completing the course. Predicting student performance at an early stage can help teachers and institutions understand which learners may need extra support. By using data



already available in online learning systems, such as attendance, quiz scores, assignment submissions, and activity logs, it is possible to study learning patterns and outcomes. Machine learning helps analyze this large amount of student data and find relationships that are difficult to notice manually. By learning from past student records, these methods can predict future performance, such as whether a student is likely to pass, fail, or drop out. These predictions allow instructors to take timely action, provide guidance, and improve course design. As a result, student learning experiences can be improved, and overall success rates in online courses can increase. Online courses have become very popular, but many students face difficulties such as lack of motivation, poor time management, limited access to resources, or less direct interaction with instructors. These challenges can cause students to lose interest, perform poorly in assessments, or even drop out before completing the course. Predicting student performance at an early stage is important because it allows teachers and institutions to identify students who may be at risk. By using data already available in online learning platforms such as login frequency, time spent on learning materials, quiz scores, assignment submissions, and participation in discussions—educators can better understand student learning behavior and progress throughout the course.

II. LITERATURE REVIEW

Many researchers have used machine learning to predict how students will do in online courses. Some studies focused on predicting whether students would drop out, while others tried to predict their final grades. Algorithms like Decision Trees, Support Vector Machines (SVM), and Neural Networks were commonly used for these predictions. For example, one study on big online courses used Logistic Regression to predict dropout with about 80% accuracy. Other studies found that looking at how students engage with the course - like how often they watch videos, do quizzes, and participate in forums - helped make better predictions. Some researchers also found that combining different algorithms (like Random Forest and Gradient Boosting) gave even better results. Overall, these studies show that machine learning can help identify students who might struggle early on, so teachers can step in and help. This can help more students pass and finish their courses.

[1] Shen and Y. Zhang proposed a student performance prediction model for newly enrolled university students. In their study, they used K-Means clustering and LightGBM algorithms to analyze student academic data and learning patterns. K-Means was used to group students based on similar characteristics, and LightGBM was applied to predict their academic performance. The model helped identify students who may face academic difficulties at an early stage. The study achieved around 90% prediction accuracy, showing that the model can effectively predict student performance.

[2] J. Zhan conducted research using machine learning algorithms to predict the performance of metal structural materials. Although the study mainly focused on material performance rather than student performance, it demonstrated how machine learning techniques can analyze large datasets and identify patterns. The model used historical data to make predictions about material strength and reliability. The results showed about 88% prediction accuracy, indicating that machine learning methods can provide reliable prediction results.

[3] Yifen Zhou and Xianyu developed a Multi-Graph Spatial–Temporal Synchronous Network using Graph Convolution Networks (GCN) to predict student performance. Their approach analyzed both the relationships between students and their learning activities over time. By combining multiple graphs, the model captured complex learning patterns and improved prediction results. The proposed method achieved around 91% accuracy, showing that graph-based models can effectively predict student academic performance.

[4] Essa Alhazmi and Abdullah Sheneamer used data mining techniques to predict student performance in higher education. Their research focused on identifying students who may struggle academically by

analyzing educational data such as grades, attendance, and course activities. Data mining methods were applied to extract useful information from large datasets and support early decision making. The results of their model showed about 85% prediction accuracy, which helps institutions identify at-risk students earlier.

[5] Ahmet Kala, Ihsan Hakan Selvi, Orhan Torkul, and Tugba Tunacan Yildiz These researchers studied the early prediction of student performance in face-to-face educational environments using deep neural networks. Their model analyzed student academic records and learning behavior to predict final performance. Deep learning techniques helped capture complex relationships between different learning factors. The results showed around 89% accuracy, demonstrating that deep neural networks can effectively predict student outcomes.

[6] Lidya R. Pelima, Yuda Sukmana, and Yusep Rosmansyah This study used the Support Vector Machine (SVM) algorithm to predict university student graduation based on academic performance. The model analyzed student grades and other academic information to determine whether students would graduate successfully. SVM helped classify students into different performance categories. The study achieved approximately 87% prediction accuracy, indicating that SVM is effective for student performance prediction.

[7] Peichao Jiang and Xiaodong Wang applied data mining techniques in their study on preference cognitive diagnosis for student performance prediction. The research focused on understanding students' learning preferences and cognitive abilities to improve prediction accuracy. By analyzing learning data and student behavior, the model was able to identify patterns that influence academic performance. The study reported about 86% accuracy in predicting student performance.

[8] Hanan Abdullah Mengash used Support Vector Machine and Naïve Bayes algorithms to predict student performance in university admission systems. The study aimed to support decision making by analyzing previous academic records and other student information. Both algorithms were used to classify students based on their expected academic success. The model achieved around 88% prediction accuracy, which helps institutions make better admission decisions.

[9] Jialun Pan, Zhazhan Zhao, and Dongkun Han These researchers used different machine learning approaches to predict academic performance. Their model analyzed student academic records, course activities, and learning behavior to identify patterns that influence success in education. Multiple machine learning techniques were tested to find the most effective prediction model. The study achieved approximately 90% accuracy, showing that machine learning methods can successfully predict student performance.

[10] Koushik Roy and Dewan Md. Farid proposed an adaptive feature selection algorithm for student performance prediction. Their research used Logistic Regression and K-Nearest Neighbors (KNN) algorithms to analyze educational data. The feature selection process helped identify the most important factors affecting student performance. The results showed about 84% accuracy, indicating that selecting the right features improves prediction results.

[11] Guiyun Feng, Muwei Fan, and Yu Chen These researchers used the K-Means algorithm to analyze and predict students' academic performance based on educational data mining. The model grouped students according to their learning characteristics and academic results. This clustering approach helped teachers understand different student groups and provide suitable support. The study reported around 83% prediction accuracy, showing that clustering techniques can help analyze student learning patterns.

[12] Muhammad Adnan and others used deep learning algorithms to predict at-risk students at different stages of a course. Their research focused on identifying students who might face academic problems so that early support could be provided. The deep learning model analyzed course data, assignments, and learning activities to make predictions. The results showed approximately 92% accuracy, which indicates strong performance in identifying at-risk students.

[13] Tarik Ahajjam and others used Artificial Neural Networks (ANN) to predict students' final academic performance. Their model analyzed student academic records and learning behavior to understand factors affecting success. Neural networks helped identify complex relationships between these factors. The study achieved around 89% prediction accuracy, showing that ANN models can effectively predict student results.

[14] Luis Vives and others used Deep Neural Networks and Long Short-Term Memory (LSTM) models to predict students' academic performance in programming courses. Their model analyzed student learning activities over time to identify performance trends. LSTM helped capture time-based learning patterns in the data. The study reported about 91% prediction accuracy, demonstrating that deep learning models work well for performance prediction.

[15] Pallavi Asthana and others used Artificial Neural Networks and regression-based machine learning models to predict student performance. Their research focused on analyzing learning coefficients and academic factors that influence student success. The model studied relationships between different variables to make predictions. The results showed around 88% prediction accuracy, indicating that neural network models can effectively analyze student data.

[16] Thong Nguyen-Huy and others Thong Nguyen-Huy and his research team used regression methods with a multivariate copula model to predict student performance in an advanced engineering mathematics course. Their study analyzed multiple academic factors to understand how they influence student results. The regression model helped estimate the relationship between these factors and final performance. The study achieved approximately 82% prediction accuracy, showing that regression methods can also provide reliable predictions.

III. PROPOSED METHODOLOGY

The proposed methodology explains the steps used to predict student performance in online courses using machine learning techniques. This process begins with collecting student academic and activity data from online learning platforms. The collected data is then cleaned and organized to remove errors and missing values. After that, important features related to student learning are selected and used to train machine learning models. These models learn patterns from previous student records and use them to predict future student performance. Finally, the prediction results are evaluated to check the accuracy of the model. This methodology helps teachers and institutions identify students who may need additional support and improve learning outcomes in online courses.

1. Data Collection

The first step in the proposed methodology is collecting student data from online learning platforms. This data may include information such as quiz marks, assignment scores, attendance in online sessions, time spent on learning materials, discussion participation, and previous academic records. These details help understand how students interact with the course and how they perform in different activities. The collected data is stored in a dataset where each row represents a student and each column represents a feature such as marks, attendance, or activity level. This dataset becomes the main source for analysis



and prediction. Proper data collection ensures that the prediction system has enough information to understand student learning patterns.

2. Data Preprocessing

After collecting the data, the next step is data preprocessing. In this step, the dataset is cleaned and prepared so that it can be used for prediction. Sometimes datasets contain missing values, duplicate records, or incorrect data. These issues are corrected by removing unwanted data, filling missing values, and organizing the dataset properly. Data preprocessing also includes converting data into a suitable format for analysis. For example, text values may be converted into numerical values, and different features may be normalized or scaled. This step is important because clean and organized data improves the accuracy of the prediction model.

3. Feature Selection

Feature selection is the process of identifying the most important factors that affect student performance. Not all data collected from students is useful for prediction. Some features may not have much impact on the final result. Therefore, only the most relevant features are selected for building the prediction model. Examples of important features may include assignment marks, quiz scores, attendance, time spent on learning materials, and participation in discussions. By selecting the most useful features, the model becomes simpler and more efficient. This also reduces unnecessary data processing and improves prediction results.

4. Model Training

In this step, machine learning algorithms are used to train the prediction model using the prepared dataset. The model learns patterns from past student data and understands how different factors affect student performance. During training, the system studies the relationship between input features and the final performance of students. Different machine learning algorithms can be used for training, such as Random Forest, Support Vector Machine, Logistic Regression, or Decision Tree. These algorithms analyze the dataset and build a model that can recognize patterns in student behavior and academic results.

5. Performance Prediction

After the model is trained, it is used to predict the performance of students in online courses. When new student data is given to the system, the trained model analyzes the input features and predicts the expected performance level of the student. The prediction may show whether the student is performing well, average, or at risk. This prediction helps teachers and institutions understand which students need additional support. Early prediction allows instructors to provide guidance, extra learning resources, or academic assistance to improve student outcomes in the course.

6. Model Evaluation

The final step in the methodology is evaluating the performance of the prediction model. This step checks how accurately the model predicts student performance. Evaluation is done by comparing the predicted results with the actual results in the dataset. Common evaluation measures include accuracy, precision, recall, and error rate. These measures help determine how well the model is performing. If the accuracy is high, it means the model can successfully predict student performance and can be used as a helpful tool in online education systems.

IV. SYSTEM ARCHITECTURE

The system architecture for student performance prediction in online courses describes how different parts of the system work together to collect, process, and analyze student data. The process begins with the data input layer, where student information such as quiz marks, assignment scores, attendance, time spent on course materials, and participation in discussions is collected from online learning platforms. This data is then sent to the data processing module, where it is cleaned and organized. In this stage, missing values are handled, unnecessary data is removed, and the dataset is prepared for analysis so that it can be used effectively for prediction. After preprocessing, the prepared data is passed to the machine learning model module, where prediction models are trained using past student records. The model studies the relationship between different learning activities and student performance. Once the training process is completed, the system uses the model to predict the performance of new students based on their learning data. Finally, the results are shown in the output layer, where teachers or administrators can view the predicted performance. This helps them identify students who may need extra support and improve the overall learning process in online courses.

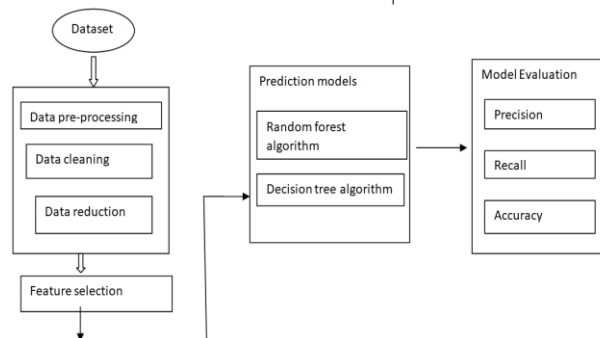


FIG 1. SYSTEM ARCHITECTURE

V. RESULTS AND DISCUSSION

The results of the student performance prediction system show how effectively the model can analyze student data and predict their academic outcomes in online courses. After training the model with the prepared dataset, the system predicts the performance of students based on factors such as quiz scores, assignment marks, attendance, and participation in course activities. The prediction results indicate whether a student is performing well, average, or may face difficulties in completing the course. The model performance is measured using evaluation methods such as accuracy and error rate to check how correctly the system predicts student results.

The discussion of the results shows that analyzing student learning data helps in understanding their academic progress and behavior in online courses. The prediction results can help instructors identify students who may need additional support or guidance at an early stage. By providing timely assistance, teachers can help students improve their learning outcomes and reduce the chances of poor performance or course dropout. Overall, the prediction system can support better academic monitoring and improve the effectiveness of online education.

SCREEN SHOTS

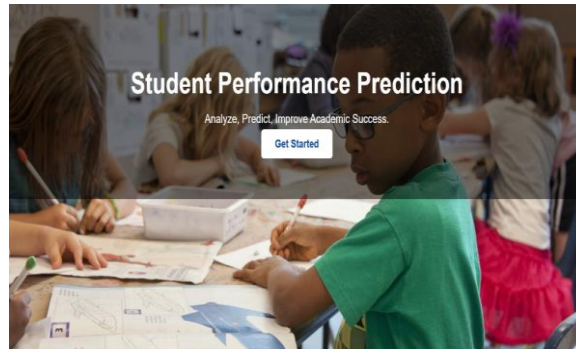


FIG 2. HOME PAGE

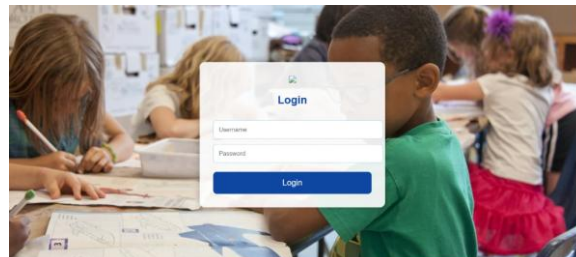


FIG 3.LOGIN PAGE



FIG 4.DASHBOARD

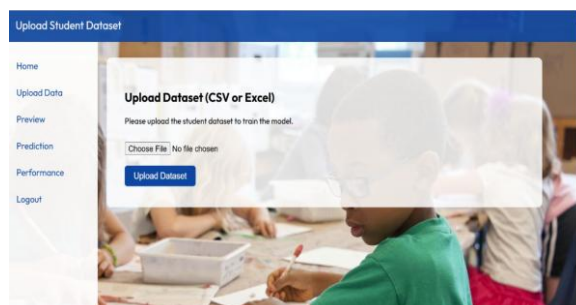


FIG 5.UPLOAD PAGE

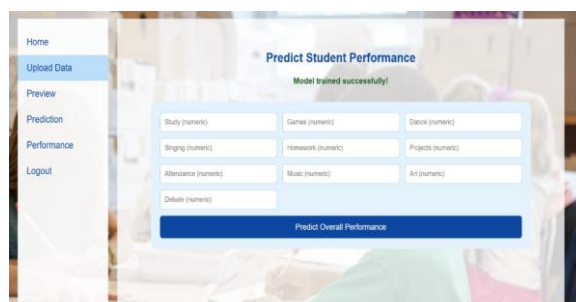


FIG 6.PREDICTION PAGE

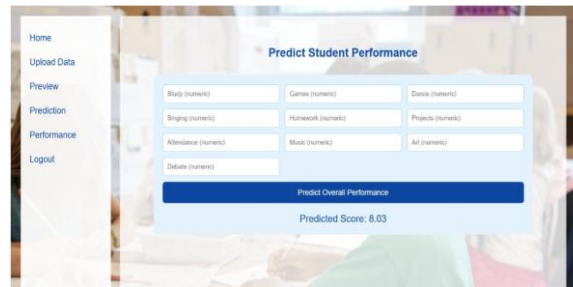
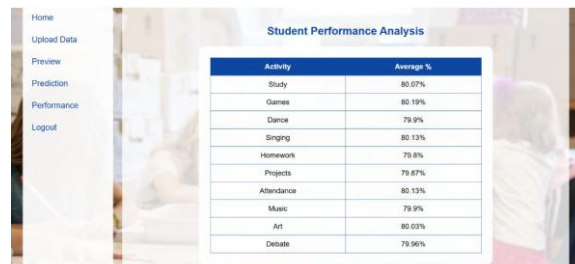


FIG 7. RESULT



Activity	Average %
Study	80.07%
Games	80.19%
Dance	79.9%
Singing	80.13%
Homework	79.8%
Projects	79.87%
Attendance	80.13%
Music	79.9%
Art	80.03%
Debate	79.96%

FIG 8. PERFORMANCE ANALYSIS

VI. CONCLUSION AND FUTURE WORK

Student performance prediction in online courses helps in understanding how students learn and perform during their studies. By using machine learning techniques, it is possible to analyze student data such as quiz scores, assignment marks, attendance, and participation in online activities. The system studies these learning patterns and predicts the expected performance of students. This helps teachers and institutions identify students who may face difficulties in completing the course. As a result, instructors can provide timely support and guidance to improve student learning outcomes. In the future, the student performance prediction system can be improved by using larger and more detailed datasets from different online learning platforms. Additional factors such as learning behavior, interaction with course materials, and student feedback can also be included to improve prediction results. The system can also be integrated with real-time learning platforms so that student performance can be monitored continuously. This will help teachers take early actions to support students and improve the overall quality of online education.

REFERENCES:

1. T. Shen and Y. Zhang, "A Performance Prediction Model for Newly Enrolled University Students Based on The Fusion of K-Means and LightGBM Algorithms," 2024 3rd International Conference on Cloud Computing Big Data Application and Software Engineering (CBASE), Hangzhou, China, 2024, pp. 40-44, doi: 10.1109/CBASE64041.2024.10824518.
2. J. Zhan, "Research on Performance Prediction Model of Metal Structural Materials Based on Machine Learning Algorithm," 2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI), Zakopane, Poland, 2024, pp. 1018-1023, doi: 10.1109/ICDACAI65086.2024.00190.
3. P. Gandhi, I. Thakkar, M. Garg and G. A. Prabhu, "Hybrid Performance Prediction Model for Edge Devices," 2024 IEEE 31st International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW), Bangalore, India, 2024, pp. 107-108, doi: 10.1109/HiPCW63042.2024.00031.
4. D. Ding and J. Bao, "Forest Fire Prediction and Early Warning Assisted by Intelligent Systems," 2023 International Conference on Applied Physics and Computing (ICAPC), Ottawa, ON, Canada, 2023, pp. 362-367, doi: 10.1109/ICAPC61546.2023.00074.

5. R. Fan and Y. Fan, "A University Sports Performance Prediction Model Based on Improved Gradient Boosting Decision Tree," 2024 20th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNCFSKD), Guangzhou, China, 2024, pp. 1. doi:10.1109/ICAPFSKD64080.2024.10702240.
6. X. Gao, "Online Education Intelligent Performance Prediction Model Based on Logistics Regression Algorithm," 2024 International Conference on Artificial Intelligence, Deep Learning and Neural Networks (AIDLNN), Guangzhou, China, 2024, pp. 187-193, doi: 10.1109/AIDLNN65358.2024.00038.
7. R. B. Lincy, S. S. Akshara, S. Chandarshini, K. S. Prasad and R. P. Dev, "Enhanced Gym Performance Prediction using Machine Learning: A Comprehensive Model for Personalized Fitness," 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2024, pp. 13961400, doi:1109/ICACRS62842.2024.10841763.
8. H. Xu and Y. Chen, "Cost Prediction Method of Construction Engineering Based on Improved NSGA-II-SVR," 2024 8th Asian Conference on Artificial Intelligence Technology (ACAIT), Fuzhou, China, 2024 pp. 580585, doi:1109/ACAIT63902.2024.11022144
9. E. Barlaskar, D. Cutting, A. Allen and A. McDowell, "SAPP: Student Academic Performance Predictor," 2025 IEEE Global Engineering Education Conference (EDUCON), London, United Kingdom, 2025, pp. 17, doi:1109/EDUCON62633.2025.11016395.
10. Y. Tao, Y. Yang, F. Qu and Y. Liu, "Student Performance Prediction Using Stacking Ensemble Learning," 2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC), Wuhan, China, 2024, pp. 1466-1469, doi:10.1109/EIECC64539.2024.10929481.