International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Paper Research

Students Performance Prediction In Online Courses Using Machine Learning Algorithms

Ms.K.Sai Durga, MCA - Assistant professor, Department of MCA, Bapatla Engineering College, Bapatla, Andhra Pradesh

Ms.EEMANA BHARGAVI, Reg No: Y25MC23019, Ms.BATHINA VENKAT RATHNA, Reg No: Y25MC23009, Mr.THATI VINOD, Reg No: Y25MC23090, Mr.PAPANABOYINA VENKATA CHAITANYA, Reg No: Y25MC23059, Department of MCA, Bapatla Engineering College, Bapatla, Andhra Pradesh.

***Abstract*—The education system is changing fast with all the new technology. This means that schools now have a lot of information about how studentsre doing in school and how they behave. The old way of checking how well students are doing like giving them tests and homework takes a time and does not always find students who are struggling early enough. This paper is about a system that uses machine learning to figure out how well engineering students will do in school by looking at lots of things about them including how well they do, in their classes and how they behave.**

The new system uses ways to figure out how students are doing like Random Forest, bagging and boosting algorithms. It helps to put student outcomes into groups like pass or fail, good, okay or bad or many other groups.

We compare these methods to see which one is the best at predicting how students will do and which one works the best.

The system automatically finds students who are struggling so we can help them sooner. Give them the support they need to learn better. The student outcomes are classified into three-class and multi-class performance categories using these methods and the system is good at detecting, at-risk students.

Additionally, the framework is designed to be scalable for large datasets and adaptable to diverse student populations. By integrating predictive analytics with a user-friendly interface and visualization tools, the proposed approach enhances decision-making for educators and institutional administrators. The experimental results demonstrate that ensemble methods outperform traditional classifiers in terms of accuracy and reliability, making the system a practical solution for modern educational performance evaluation.

Keywords: 1. Student performance prediction, 2. machine learning, 3. educational data mining, 4. learning analytics, 5. online learning systems, 6. ensemble learning.

I. INTRODUCTION

Students are the people who go to schools and colleges. The grades they get and what they learn have an impact on the country. They grow up to be workers people who come up with new ideas and business owners. This helps the country do better socially and economically.

Nowadays a lot of students learn on computers and the internet. This means that schools and colleges have a lot of information about what students do and how they learn. They have data on things like grades and how students interact with each other. This has led to people studying Educational Data Mining and Learning Analytics. These are fields where people try to understand what all this data means. They want tofind information, in all the data that schools and colleges have.

People use data mining and machine learning techniques to find patterns and trends in data. They want to make decisions based on this information.

The internet has changed the way we learn with things like classes and Learning Management Systems. We also have Massive Open Online Courses. All of these things give us a lot of information about learners, like how they attend classes how well they do on tests and how much they participate in online activities. Data mining and machine learning techniques are used to understand this learner data. We have kinds of data sources. These data sources can be put into two groups. The first group is centralized systems. This includes things like LMS platforms. The second group is decentralized sources. This includes online course portals and web-based learning tools, like online course portals and web-based learning tools

Higher education institutions have a lot of trouble keeping track of how studentsre doing and finding out which students are struggling with their school work. The old ways of checking how students are doing which mostly involve teachers giving tests and checking

homework by hand do not work well. These old ways do not give us warning signs when a student is doing poorly. So people really want to have computer systems that can tell us how a student will do in school. They want these systems to be accurate and work quickly. Higher education institutions need these systems to help students who're, at risk of not doing well in school.

Machine learning has gotten a lot better lately. This means we can make models that try to figure out how well a student will do in school. These models look at a lot of things like what the student's studying and how they behave. They use math techniques like classification algorithms and ensemble learning methods. These techniques are really good at making sure the predictions are correct.

This is helpful because it means we can do something to help the student before they struggle. We can make a plan that's just for them and help them learn in a way that works best for them. Machine learning and these models support this kind of planning. Make it easier to make good decisions, about what to do.

This paper focuses on the application of machine learning algorithms for predicting student performance in online courses. By utilizing educational data collected from digital learning platforms, the proposed system aims to classify students into different performance categories and identify those at risk of academic failure. The integration of predictive analytics within online education environments can enhance instructional effectiveness, improve student retention, and support educators in making timely and informed decisions.

II. LITERATURE SURVEY

Recent advancements in machine learning and data ana- lytics have enabled intelligent systems across various domains, including healthcare, smart environments, business processes, and education. Bashar et al. [1] proposed an optimized deep learning approach using transfer learning models such as ResNet and EfficientNet for COVID-19 pneumonia detection, demonstrating improved classification accuracy through fine- tuning techniques. Similarly, Latif et al. [9] applied deep neural network-based feature extraction combined with Support Vec- tor Machines (SVM) for glioma tumor classification, achieving high performance and validating the effectiveness of hybrid machine learning approaches.

Machine learning has also been utilized in smart systems, as shown by Mahmoud et al. [2], who developed an infant sound classification system using novel acoustic features and SVM classifiers. The system effectively distinguished between various infant sounds, contributing to smart city and home monitoring solutions. Furthermore, Latif et al. [11] introduced a drone-based deep learning system for plant disease detec- tion and automated spraying, highlighting real-time intelligent decision-making in precision agriculture.

In the educational domain, significant research has focused on predicting student performance using machine learning techniques. Latif et al. [3] proposed an AI-based prediction system employing Decision Trees, Random Forests, and SVM to identify at-risk students based on academic features such as attendance and previous grades. The system achieved high prediction accuracy and supported early academic interven- tions. Similarly, Bujang et al. [4] developed a multi-class grade prediction model using ensemble learning techniques including Random Forest and XGBoost, which significantly enhanced classification performance across multiple student outcome categories.

Gupta et al. [5] introduced an ensemble model combining bagging and boosting techniques for early identification of at- risk students in higher education. Their approach outperformed traditional classifiers, providing an effective early warning sys- tem. Kostopoulos et al. [6] explored semi-supervised learning techniques in distance education environments to handle lim- ited labeled data, demonstrating accurate prediction of student performance using Decision Trees and KNN classifiers.

Ma and Zhou [7] applied optimized SVM and Decision Tree models for predicting student pass rates, achieving improved accuracy through hyperparameter tuning and feature selection. Additionally, Sa´iz-Manzanares et al. [8] utilized educational data mining techniques on Learning Management System (LMS) data to monitor student engagement and predict learning outcomes using Random Forests and neural networks. Their system enabled instructors to identify struggling students through behavioral and interaction data.

Beyond education, machine learning has been applied in business optimization, as demonstrated by Al-Anqoudi et al. [10], who proposed a machine learning-based framework for Business Process Re-Engineering to improve operational efficiency. Nti et al. [12] conducted a systematic review of machine learning techniques in stock market prediction, concluding that ensemble models and neural networks signif- icantly enhance predictive accuracy across financial datasets. Overall, existing studies highlight the effectiveness of ma- chine learning and ensemble techniques in predictive analytics across multiple domains. However, there remains a need for integrated, scalable models specifically designed for online learning environments that leverage both academic and be- havioral data to accurately predict student performance and identify at-risk learners. This research aims to address this gap by implementing and evaluating machine learning algorithms

for performance prediction in online courses.

III. METHODOLOGY

.

The new system is supposed to figure out how well students will do in courses. It uses computer programs to look at the information about what students do and how they learn.

The system has a main steps: we collect information about students we make sure the information is good to use we pick the important parts we teach the computer program we see how well the program works and then we look at the results to see what they mean for student performance, in online courses.

A. Data Collection:

The Learning Management Systems or LMS is where we get the dataset from. This Learning Management Systems has a lot of information about the students. It includes things like the student details their attendance records, the scores they get on assignments the results of their quizzes how much they participate in discussion forums and how well they do in the course overall. These things, like the assignment scores and the quiz results, show what the students have achieved in school.. Things like attendance records and participation in discussion forums show how the students behave when they learn. We need to know all these things about the students, including the assignment scores and the quiz results and the attendance records and the participation in discussion forums to make predictions, about the students and the Learning Management Systems.

B. Data Preprocessing:

Raw educational data usually has some problems. It can have missing information, mistakes and things that do not make sense. So we use some techniques to make the data better. We fill in the missing information with the values and we remove the things that are not important or that are repeated. We also change the data so that it is in a format that the machine learning models can understand. We do this by making sure all the information is in numbers because the models work best with numbers. Raw educational data needs this work so that it is good enough, for the machine learning models.

C. Feature Selection:

We look at the things that really affect how well students do in school. We use math techniques to figure out what is important. This helps us simplify things and make our model work better. We focus on the behavioral things that have the biggest impact, on student performance. We want to know what really matters for student performance.

D. Model Development:

To figure out how well students will do we use different kinds of machine learning classification algorithms. We use Decision Trees, Support Vector Machines, k-Nearest Neighbors and Random Forest to make predictions about student performance. We also use learning techniques like bagging and boosting to make our predictions more accurate and to make our models stronger.

The machine learning classification algorithms are trained using student data. The student performance levels are then divided into a groups. Some are just two groups some have three groups and some have many groups. The machine learning classification algorithms help us predict which group each student will be, in.

E. Performance Evaluation:

The models we have trained are checked to see how well they work by using things like accuracy and precision and recall and F1-score and confusion matrix.

We use cross-validation to make sure the results are real and to stop the models from getting too good at one thing.

The performance of the models is compared to the performance of individual classifiers to figure out what works best with the models.

F. System Implementation:

A user-friendly interface is developed to display prediction results and performance insights for educators and administrators. The system provides early warnings for at-risk students, enabling timely academic intervention and personalized learning support.

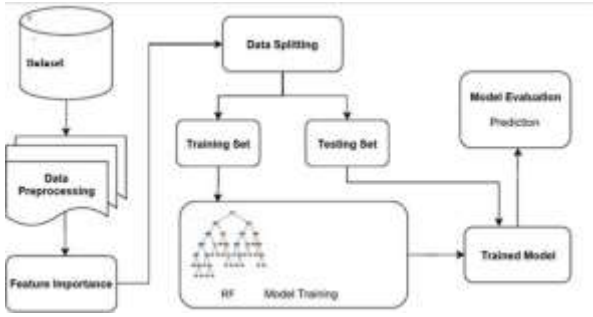


Fig. 1. Architecture diagram

IV. ALGORITHM

Algorithm for Student Performance Prediction:

The proposed system follows a structured machine learning workflow to predict student performance in online courses. The algorithm integrates data

preprocessing, feature selection, model training, and evaluation to generate accurate predictions.

Student Performance Prediction Using Machine Learning Input: Student dataset containing academic records, atten- dance, online activity logs, and assessment scores

Output: Predicted performance category of students (e.g., Pass/Fail or Multi-class levels)

Step 1: Load the student performance dataset from the online learning platform.

Step 2: Perform data preprocessing: a) Handle missing values using imputation techniques
b) Remove irrelevant and noisy attributes
c) Encode categorical variables into numerical format
d) Normalize numerical features

Step 3: Select significant features using correlation analysis or feature importance methods.

Step 4: Split the dataset into training and testing sets.

Step 5: Train multiple machine learning classifiers: a)
Decision Tree
b) Support Vector Machine (SVM)
c) k-Nearest Neighbours (KNN)
d) Random Forest

Step 6: Apply ensemble learning techniques such as bagging and boosting to improve prediction performance.

Step 7: Evaluate trained models using performance metrics:
a) Accuracy
b) Precision
c) Recall
d) F1-score

Step 8: Select the best-performing model based on evaluation results.

Step 9: Predict student performance categories for new or unseen data.

Step 10: Display results through a user-friendly interface for educators.

Algorithm Description:

The algorithm begins by collecting and cleaning student data from online learning platforms. Relevant academic and behavioral features are selected to improve prediction accuracy. Multiple classification models are trained and compared to identify the most effective approach. Ensemble methods are employed to enhance robustness and minimize overfitting. The final model predicts student performance levels, enabling early identification of at-risk learners and supporting timely academic interventions.

V. SYSTEM ANALYSIS

The online education platforms are growing fast and this is creating a lot of data about students. This data includes things, like records and scores and how students interact with the online system. This information can be really useful to help students learn better.. The way we check how students are doing now is pretty old fashioned. We mostly do it by hand. We only check in with students every now and then. This takes a lot of time. We do not always know how students are doing right now. Sometimes we do not even notice when a student is struggling until it is too late. Online education platforms can help us with this.

A. Existing System :

In school student performance is looked at mostly by doing quizzes, assignments, mid-term tests and final exams. Teachers or school staff usually go over how studentsre doing by hand. This way of doing things has some problems:

Limited ability to analyze large datasets

effectively Delayed identification of at-risk

students

Lack of personalized learning support

High dependency on human effort and subjective

judgment Inability to utilize behavioral and

engagement data from
online platforms

Online education is getting bigger and bigger. The old systems we have been using are not good enough to deal

with all the huge amounts of educational data that come with it. Online education needs something to handle all this data.

B. Proposed System:

The new system is a way to guess how well students will do in classes. It looks at things like grades. How students behave when they use online learning tools like Learning Management Systems and other digital learning platforms. The system gets all the information together picks out the parts and uses special methods

to make predictions. Then it can tell us how well students will probably do in their classes. The system uses student data to make these guesses. It is really good at figuring out what students will achieve. The system is, about predicting student performance in online courses.

We use kinds of machine learning classifiers like Decision Trees and Support Vector Machines and k-Nearest Neighbors and Random Forest to make our predictions better.

We also use learning techniques like bagging and boosting to make the system stronger and more reliable.

The machine learning classifiers we use are really important because they help us with prediction accuracy.

Decision Trees and Support Vector Machines and k-Nearest Neighbors and Random Forest are all machine learning classifiers that we use to improve our predictions.

The ensemble learning techniques like bagging and boosting make our system better. We get more accurate predictions, from the machine learning classifiers.

The system they are talking about can watch how students are doing all the time. This means teachers can find out which students are having trouble on. Teachers can use this information to help these students away with special plans that are just for them. The system is really good, for students because it helps teachers give them the help they need when they need it. The student progress is what is important here. The system helps with that.

C. Advantages of the Proposed System: Automated and efficient performance prediction

Improved accuracy through machine learning and ensemble techniques

Early identification of struggling students

Utilization of both academic and behavioral

data Scalable for large and diverse student

datasets

Supports data-driven decision making for educators

VI. RESULT ANALYSIS

The student performance prediction system was tested to see how well it works. We used kinds of machine learning to do this like Decision Tree, Support Vector Machine, k-Nearest Neighbors and Random Forest. We also tried using learning techniques like bagging and boosting to make the system better, at predicting student performance. The student performance prediction system was trained on one set of data. Tested on another set. We did this times using a method called cross-validation to make sure the results of the student performance prediction system are trustworthy.

The things we look at to see how well something works are accuracy, precision, recall and F1-score. These things help us understand how good the classification is. Random Forest was the best at getting things right it did better than Decision Tree, SVM and KNN models. This is because Random Forest uses decision trees together which helps it not get too good at one thing and bad, at others and it gets better at working with new things. Random Forest is really good because of the way it combines decision trees, which is why it does a good job.

The ensemble learning models did a good job of predicting things.

Bagging techniques made the models more stable.

Boosting methods were also helpful because they paid attention to the things that the models got wrong.

The ensemble learning models were better at getting things right than classifiers.

This shows that ensemble learning models are very good, at dealing with educational data.

The ensemble learning models are really effective.

When we try to figure out how well students will do in a class with different levels the Random Forest and boosted models do a great job of predicting this. They work well for all levels. We looked at the results in a table called a confusion matrix. This showed that the Random Forest and boosted models were able to put students into the right category of how well they will do. The Random Forest and boosted models did not often get it wrong when trying to decide between two levels that're close together. The Random Forest and boosted models are really good, at predicting how well students will do in classes.

The new system is really good at finding students who're at risk of falling behind. It does a job of catching

these students, which is important so we can help them early on.

The results show that using information about what students do, in school and how they behave really helps us figure out which students might need help when they are learning online.

Overall, the experimental results demonstrate that machine learning algorithms, particularly ensemble methods, provide reliable and accurate predictions of student performance in online courses. The proposed approach offers a practical solution for educational institutions to monitor student progress, support personalized learning, and enhance academic outcomes.



Fig : 2. Main Page



Fig . 3 . Starting Prediction



Fig. 4. Withdrawn Prediction



Fig 5. Evaluation

VII. CONCLUSION

This paper is about using machine learning to predict how students will do in courses. It looks at the behavioral data from digital learning platforms. Online education is growing fast. This means we need systems that can handle a lot of data and give us useful information, about how students are learning. The old ways of checking how students are doing often do not find students who are struggling until it is too late. This makes it hard to help these students when they need it. Machine learning can help with this by looking at student performance and behavioral data from courses.

The new system uses kinds of machine learning classifiers like Decision Tree, Support Vector Machine, k-Nearest Neighbors and Random Forest. It also uses learning techniques like bagging and boosting. When we tried out the system we found out that the based models, especially Random Forest were really good at making predictions. They were better at it than the classifiers. The system looks at lots of features including features and features that show how engaged someone is. This makes the system really good at predicting how someone will perform. The features from academics and engagement really help make the predictions more reliable. The machine learning classifiers, including Decision Tree, Support Vector Machine, k-Nearest Neighbors and Random Forest are important, for the system.

The system did a job of finding students who might have trouble with their school work. This means teachers can help these students on and give them learning plans that are just right for them. The system helps students stay in school learn more and makes sure teachers have the information they need to make decisions about how to teach the students. The system is really good, at identifying students who need help. This helps the students get the support they need to do well in school.

In future work, the system can be extended by incorporating deep learning models, real-time learning analytics, and larger datasets from diverse educational platforms.

Additionally, integrating adaptive feedback mechanisms and recommendation systems may further enhance personalized learning experiences. Overall, the proposed approach demonstrates the effectiveness of machine learning algorithms in supporting modern online education systems and improving student performance evaluation.

REFERENCES

[1]A. Bashar, G. Latif, G. Ben Brahim, N. Mohammad, and J. Alghazo, “COVID-19 pneumonia detection using optimized deep learning tech niques,” Diagnostics, vol. 11, no. 11, p.
1972, Oct. 2021, doi: 10.3390/diagnostics11111972.

[2] Rongali, L. P. (2025). The Trinity of Cybersecurity: DevSecOps, Cloud Security and Cryptography in The Digital Age. SSRN Electronic Journal.

[3] Nandigama, N. C. (2019). Energy-Efficient Virtual Machine Placement in Cloud Data Centers Using Reinforcement Learning-Enhanced Adaptive Greedy Dingo Optimization Algorithm. Research Journal of Nanoscience and Engineering, 3(1), 25–32. https://doi.org/10.22259/2637-5591.0301006.

[4]S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, and N. A. Md. Ghani, "Multiclass prediction model for stu dent grade prediction using machine learning," IEEE Access, vol. 9, pp. 95608–95621, 2021.

[5] Nandigama, N. C. (2018). Advanced Deep Learning Architecture for Multimodal Biometric Authentication Using Feature-Level Fusion and Dimensionality Reduction. Research Journal of Nanoscience and Engineering, 2(3), 34–38. https://doi.org/10.22259/2637-5591.0203005.

[6] Todupunuri, A. (2023). The Role of Artificial Intelligence in Enhancing Cybersecurity Measures in Online Banking Using AI. International Journal of Enhanced Research in Management & Computer Applications, 12(01), 103–108. https://doi.org/10.55948/ijermca.2023.01015

[7] Rongali, L. P. (2025). Green DevOps Metrics for Utility Operations. https://doi.org/10.36227/techrxiv.175433211.13655773/v1

[8]M. C. Sa´iz-Manzanares, J. J. Rodr´ıguez-D´ıez, J. F. D´ıez- Pastor, S. Rodr´ıguez-Arribas, R. Marticorena-Sa´nchez, and
Y. P. Ji, "Monitoring of student learning in learning management systems: An application of educational data mining techniques," Appl. Sci., vol. 11, no. 6, p. 2677, Mar. 2021, doi: 10.3390/app11062677.

[9] Todupunuri, A. (2024). Develop Machine Learning Models to Predict Customer Lifetime Value for Banking Customers, Helping Banks Optimize Services. International Journal of All Research Education & Scientific Methods, 12(10), 1254–1259. https://doi.org/10.56025/ijaresm.2024.1210241254.

[10] Y. Al-Anqoudi, A. Al-Hamdani, M. Al-Badawi, and R. Hedjam, "Using machine learning in business process re-engineering," Big Data Cognit. Comput., vol. 5, no. 4, p. 61, Nov. 2021, doi: 10.3390/bdcc5040061.

[11] G. Latif, J. Alghazo, R. Maheswar, V. Vijayakumar, and M. Butt, "Deep learning based intelligence cognitive vision drone for automatic plant diseases identification and spraying," J. Intell. Fuzzy Syst., vol. 39, no. 6, pp. 8103–8114, Dec. 2020, doi: 10.3233/jifs-189132. [12] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," Artif. Intell. Rev., vol. 53, no. 4, pp. 3007–3057,Aug.2019,do