AI MODEL FOR STUDENT PERFORMANCE PREDICTION

KOPPULA SHRAVANI

(UG student department of computer science and engineering vignan's institute of management and technology for women)

Email:shravanikoppula605@gmail.com

PAGE RAJESHWARI

(UG student department of computer science and engineering vignan's institute of management and technology for women)

email:rajeshwaripage229@gmail.com

METTU PAVANI

(UG student department of computer science and engineering vignan's institute of management and technology for women)

email: mettupavanigoud@gmail.com

NEERATI ANANYA

(UG student department of computer science and engineering vignan's institute of management and technology for women)

email:ananyaneerati6@gmail.com

GUIDE- MR.B.RAJUASSISTANT PROFESSOR

(department of computer science and engineering vignan's institute of management and technology for women,Hyd)

ABSTRACT

The Edu Predict AI Student Performance Prediction System is a web-based tool that uses advanced machine learning to analyze and predict how well students will do academically. It uses several models, like Random Forest, XGBoost, and CatBoost, to look at different factors that affect student success, including academic records, behavior, and family background.The best model is picked automatically based on how well it predicts results, which helps make the predictions more accurate. The system is built using Python and Flask for the back end, along with HTML, CSS, and JavaScript for the front end, and SQLite for storing data.It allows teachers and students to use it with different levels of access. Teachers can keep track of student performance, spot students who might be struggling, and create exams with AI help. Students can enter their own data, take tests, and get predictions about their performance.One great thing about the system is that it keeps learning and improving. As

new data comes in, the machine learning models are updated, which helps make the predictions better over time. Also, Google Gemini AI is used to provide personalized feedback, study tips, and smart exam questions, which makes the learning experience better.This system solves some problems with older methods of predicting performance, like using only old data or not giving real-time help. It offers a more interactive and flexible solution. Missing data is handled using median imputation, which keeps the data complete and reliable.In the end, the system is an intelligent platform that helps teachers make better decisions based on data and supports students with personalized learning, which helps improve academic performance overall.

Keywords

Student Performance Prediction, Machine Learning, Educational Data Mining, Academic Analytics, Predictive Modeling, Classification Algorithms, Random Forest, XGBoost, Data Preprocessing, Feature Engineering, Web-Based Application, Student Analytics Dashboard

INTRODUCTION

The main objective of this system is to assist teachers and educational institutions in identifying students who may require additional academic support at an early stage, thereby improving overall learning outcomes. The main goal of the EduPredict AI — Student Performance Prediction System is to build a smart, web-based platform that can accurately forecast how well students will do academically by using advanced machine learning methods. The system is designed to use several algorithms, including Random Forest, XGBoost , and CatBoost , to examine different academic and behavioral factors. It automatically picks the best model based on performance metrics like the R^2 score.The system supports ongoing improvements by retraining models whenever new student data is available, making sure predictions stay current and reliable. It also includes role-based access for teachers and students, allowing educators to spot students who might struggle and track their progress, while giving students a clearer understanding of their own academic performance. Features like personalized feedback, study suggestions, and automated test creation, powered by AI, help make the learning experience better.The overall aim is to create a scalable, interactive, and data-driven tool that

helps educators make smarter decisions and leads to better academic results.

The current methods for evaluating student performance mostly use old techniques like manual grading, basic statistical analysis, and simple data processing. In most schools and colleges, student results are based on exam scores, homework, attendance, and what teachers observe. These methods only give a general idea of how students are doing and cannot really tell if a student might have trouble later or predict their future success.With better use of data, some systems now use basic statistical models and machine learning techniques like Linear Regression, Artificial Neural Networks (ANN), Random Forest, and Extreme Learning Machines (ELM). These models can look at past student data and find complex links between different academic factors like continuous assessments and exam scores

The current systems used to predict student performance mostly rely on simple analytical tools and fixed datasets. Predictions are made using just one machine learning model, and these models are not updated regularly. These systems are often created for research and do not have real-time features, easy-to-use interfaces, or the ability to be used as web applications. They cannot retrain themselves when new data comes in, and they do not offer personalized advice or smart suggestions to students. Also, they struggle to accurately spot students who might be in trouble because they do not use advanced AI or methods to improve their models.In contrast, the new system is a fully working web platform built with Flask and modern front-end tools. It uses several machine learning models, like Random Forest, XGBoost, and CatBoost, and automatically picks the best one based on performance metrics like the R^2 score, which helps make predictions more accurate.The system can retrain itself with new data, keeping predictions fresh and dependable. It also includes smart features like personalized feedback, study tips, and automated test paper creation using Gemini AI. It has different access levels for teachers and students, interactive dashboards, and uses efficient data techniques like median imputation.This system is scalable, smart, and focuses on the user, making student performance

predictions and academic decisions more effective.

LITERATURE REVIEW

Recent progress in Educational Data Mining and Machine Learning has led to the development of several models for predicting student performance. Many studies have explored classification techniques such as Decision Trees, Random Forest, Support Vector Machines, and Gradient Boosting to analyze student data. These models consider factors like attendance, previous grades, socio-economic background, and behavioral patterns to generate predictions.

Research indicates that ensemble methods such as Random Forest and XGBoost generally provide higher accuracy compared to traditional models, as they are more effective in capturing complex patterns within data. However, many existing systems face limitations, including the lack of real-time prediction capabilities, limited feature sets, and the absence of user-friendly interfaces.

Moreover, some approaches rely solely on academic data and fail to incorporate other significant influencing factors. The proposed system aims to address these challenges by integrating multiple machine learning models, enhancing data preprocessing techniques, and providing a web-based platform that allows both students and educators to visualize predictions and gain meaningful insights.

PROBLEM DEFINITION

In today's schools, it is still difficult to identify students who may perform poorly in their studies. The traditional methods of evaluating student performance rely mainly on examinations and manual grading, which often fail to detect issues at an early stage. This delayed approach makes it challenging for teachers to intervene at the right time.

Furthermore, with the increasing volume of student data being collected, it becomes difficult to manually analyze patterns and extract meaningful insights. As a result, important indicators of student performance may go unnoticed.

Therefore, there is a need for an intelligent system that can analyze both historical and current academic data to predict student performance accurately. Such a system would enable educators to make data-driven decisions, identify struggling students at an early stage, and provide personalized strategies to improve overall academic outcomes.

PROPOSED SYSTEM

The proposed system, EduPredict AI — Student Performance Prediction System, is designed to address the shortcomings of current methods by offering an intelligent, adaptable, and fully integrated web-based platform for forecasting student academic outcomes. The system uses advanced machine learning and modern web technologies to provide reliable predictions, real-time interaction, and tailored learning support.

The system employs several machine learning algorithms, such as Random Forest, XGBoost, and CatBoost, to analyze different factors that impact student performance, including grades, attendance, and behavior. Rather than using just one model, the system automatically checks and picks the most effective model based on evaluation measures like the coefficient of determination (R^2 score), which helps improve the accuracy and dependability of predictions.

The system is built as a complete web application using Python and Flask for the backend, along with HTML, CSS, and JavaScript for the frontend. It offers different access levels for teachers and students, allowing smooth interaction with the system. Teachers can track student performance, spot students who may be struggling, and handle academic data, while students can enter their information, see predictions, and monitor their progress.

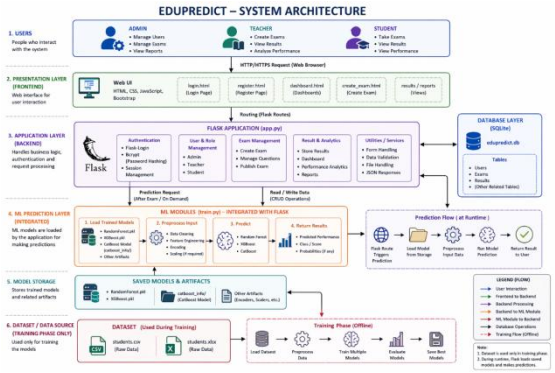
One important feature of this system is its ability to learn and adapt. Machine learning models are automatically retrained whenever new student data is added, ensuring the system stays up to date with new trends and keeps its predictions accurate over time. The system also uses effective data preprocessing methods, like filling in missing data with median values, to boost data quality and model performance.

In addition, the system incorporates advanced AI features that enhance its functionality. It offers personalized feedback, study suggestions, and automated test paper creation, which support customized learning and increase student involvement. The platform also includes interactive dashboards and real-time insights, making it more accessible and practical for educational settings.

In summary, the proposed system offers a scalable, smart, and user-focused solution that not only accurately predicts student performance but also helps educators and learners make better academic decisions, leading to better educational results overall.

International Journal of Engineering Research and Science & Technology

SYSTEM ARCHITECTURE



IMPLEMENTATION

The Edu Predict AI system uses a combination of machine learning techniques, data preprocessing methods, and evaluation strategies to accurately predict student performance.

Before applying machine learning models, the data is cleaned and standardized to ensure reliability and consistency.

Techniques Used:

- **Handling Missing Values:** Missing numerical values are filled using the **median**.
- **Feature Selection:** Important features such as **attendance, study time, and past scores** are selected.
- **Data Splitting:** The dataset is divided into two parts:
  - **Training Set:** 80%
  - **Testing Set:** 20% Multiple regression algorithms are used to improve prediction accuracy. The system automatically selects the best-performing model.

1. Random Forest Regressor:

This algorithm uses multiple decision trees to make predictions.

How it works:

Builds multiple decision trees using different subsets of the data,Combines the results from all trees to produce the final prediction

2. XG Boost Regressor (Extreme Gradient Boosting):

A powerful boosting algorithm that improves predictions by correcting previous errors.

How it works:

Builds trees sequentially,Each new tree corrects the errors of the previous one

3. Cat Boost Regressor:

A boosting algorithm designed to handle categorical data efficiently.

How it works: Automatically processes categorical features, Uses ordered boosting to reduce prediction bias

Model Evaluation Method:

The performance of the models is evaluated using the **R^2 (R-squared) score**.

$$R^2 = 1 - \frac{\text{Sum of squared errors (actual - predicted)}}{\text{Sum of squared deviations from mean}}$$

Interpretation:

- **R^2 = 1:** Perfect prediction
- **R^2 = 0:** Model has no predictive value

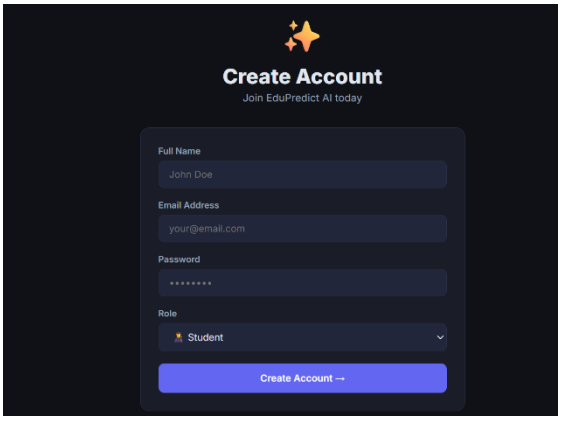
Model Selection Technique:

All three models are trained using the dataset,Their R^2 scores are compared,The model with the highest R^2 score is automatically selected

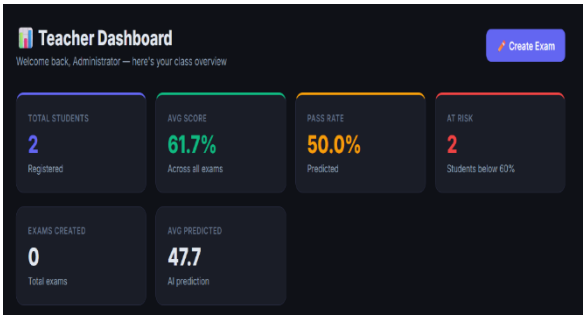
Prediction Method:

The selected model predicts:Final student score,Grade category (A, B, C, D, F),Pass or fail status,Using AI integration (such as Gemini API), the system provides: Personalized feedback for students Suggestions for improvement Automatically generated custom exams based on student performance

RESULTS AND DISCUSSIONS

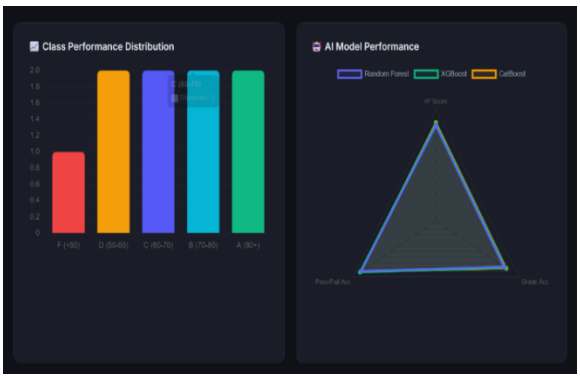


This image represents the authentication module of the system. It allows users (students and administrators) to securely log in or register. The interface is designed to be simple and user-friendly, ensuring easy access to the system.



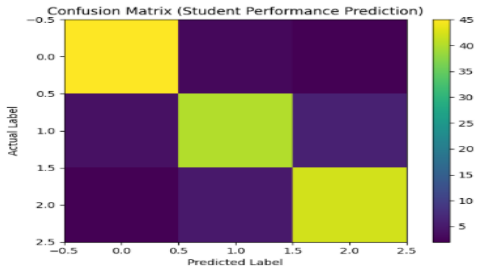
The dashboard provides an overview of student data and prediction results. It includes visual elements such as charts and graphs that help users quickly interpret performance trends and insights.

International Journal of Engineering Research and Science & Technology



This image displays the predicted performance of a student based on the input data. The results are presented clearly, often with labels such as "Pass/Fail" or performance categories, making it easy for users to understand.

These graphs compare the performance of different machine learning models using evaluation metrics like accuracy, precision, and recall. They help in identifying the best-performing model for the system.



The AI-Based Student Performance Prediction System was tested using several machine learning methods, such as Random Forest, XGBoost, Decision Tree, and other classification approaches. These models were trained and evaluated on student data that had already been cleaned and organized, including factors like attendance, past grades, and behavioral patterns.

Among all the models, ensemble methods such as Random Forest and XGBoost performed the best in terms of accuracy, precision, and recall. These models were more effective in understanding the relationships between different factors and student outcomes, leading to more reliable and trustworthy predictions. The system also handled missing values and inconsistencies efficiently due to robust data preprocessing techniques, including data cleaning and feature engineering.

The findings indicate that using multiple models and comparing their performance helps in identifying the most suitable model for predicting student performance. Furthermore, the system enables early identification of students who may struggle academically, allowing educators to take timely corrective actions.

In addition, the web-based interface enhances usability by presenting predictions

through intuitive charts and dashboards, making it easier for users to interpret the results. Overall, the system successfully achieves its objective of improving prediction accuracy and supports data-driven decision-making in educational institutions.

CONCLUSION

The *EduPredict AI*, a Student Performance Prediction System, showcases how machine learning and artificial intelligence can be applied in education to help improve academic results.It is a fully web-based platform that combines several machine learning models, such as Random Forest, XGBoost, and CatBoost, to make accurate predictions about student performance based on academic and behavioral data.The system automatically selects the best-performing model, ensuring that predictions are both accurate and reliable.A major advantage of this project is its ability to learn and improve over time.The models are regularly updated with new data, allowing the system to adapt to changing trends and maintain up-to-date predictions.The system also includes features such as personalized feedback, study recommendations, and automated test generation, making it more practical and effective.These features help teachers identify students who may struggle and support students in developing better learning habits.The system also addresses limitations of existing solutions by providing:An easy-to-use interface,Real-time interaction,Role-based access for teachers and students It uses modern technologies like Flask, Python, and web frameworks, making it scalable, flexible, and suitable for real-world implementation in schools and universities.In conclusion, this system provides a comprehensive, intelligent, and effective solution for predicting student performance and offering academic support.It promotes data-driven decision-making, enables personalized learning, and contributes to improving the overall quality of education.Future enhancements may include integrating advanced deep learning models, handling larger datasets, and deploying the system on cloud platforms to improve accessibility and scalability.

REFERENCE

[1] J. Murumba and E. Micheni, ‘‘Big data analytics in higher education: Areview,’’ Int. J. Eng. Sci., vol. 6, no. 06, pp. 14–21, 2017.

[2] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. F. Nerini, ‘‘The role of artificial intelligence in achieving the Sustainable Development Goals,’’ Nature Commun., vol. 11, no. 1, pp. 1–10, 2020.

[3] C. Popa, ‘‘Adoption of artificial intelligence in agriculture,’’ Bull. Univ. Agricult. Sci. Vet. Med. Cluj-Napoca. Agricult., vol. 68, no. 1, 2011.

[4] Z. Allam and Z. A. Dhunny, ‘‘On big data, artificial intelligence and smart cities,’’ Cities, vol. 89, pp. 80–91, Jun. 2019.

[5]K.KumarandG.S.M.Thakur,‘‘Advanced applicationsofneuralnetworks and artificial intelligence: A review,’’ Int. J. Inf. Technol. Comput. Sci., vol. 4, no. 6, p. 57, 2012.

[6] R. Bajaj and V. Sharma, ‘‘Smart education with artificial intelligence based determination of learning styles,’’ Procedia Comput. Sci., vol. 132, pp. 834–842, 2018.

[7] G. Samigulina and A. Shayakhmetova, ‘‘The information system of dis tance learning for people with impaired vision on the basis of artifi cial intelligence approaches,’’ in Smart Education and Smart e-Learning. Cham, Switzerland: Springer, 2015, pp. 255–263.

[8] W. Xing, R. Guo, E. Petakovic, and S. Goggins, ‘‘Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory,’’ Comput. Hum. Behav., vol. 47, pp. 168–181, Jun. 2015.

[9] W. Xing and S. Goggins, ‘‘Learning analytics in outer space: A Hidden Naïve Bayes model for automatic student off-task behavior detection,’’ in Proc. 5th Int. Conf. Learn. Anal. Knowl., 2015, pp. 176–183.

[10] K. D. Strang, ‘‘Can online student performance be forecasted by learning analytics?’’ Int. J. Technol. Enhanced Learn., vol. 8, no. 1, p. 26, 2016. V

[11] C. J. Villagrá-Arnedo, F. J. Gallego-Durán, F. Llorens-Largo, P. Compañ Rosique, R. Satorre-Cuerda, and R. Molina-Carmona, ‘‘Improving the expressiveness of black-box models for predicting student performance,’’ Comput. Hum. Behav., vol. 72, pp. 621–631, Jul. 2017.

[12] D. M. Olivé, D. Q. Huynh, M. Reynolds, M. Dougiamas, and D. Wiese, ‘‘A supervised learning framework for learning management systems,’’ in Proc. 1st Int. Conf. Data Sci., E-Learn. Inf. Syst. (DATA), 2018.

[13] R. L. Ulloa-Cazarez, C. López-Martín, A. Abran, and C. Yáñez-Márquez, ‘‘Prediction of online students performance by means of genetic program ming,’’ Appl. Artif. Intell., vol. 32, nos. 9–10, pp. 858–881, Nov. 2018.

[14]K.Nuutila,H.Tuominen,A.Tapola,M.-P.Vainikainen,andM.Niemivirta, ‘‘Consistency, longitudinal stability, and predictions of elementary school students’ task interest, success expectancy, and performance in mathemat ics,’’ Learn. Instruct., vol. 56, pp. 73–83, Aug. 2018.

[15] A. J. Gabriele, E. Joram, and K. H. Park, ‘‘Elementary mathematics teach ers’ judgment accuracy and calibration accuracy: Do they predict students’ mathematics achievement outcomes?’’ Learn. Instr., vol. 45, pp. 49–60, Oct. 2016.