DOI:10.22413/ijatest/2026/V11/i2/01

Adaptive Student Performance Prediction and Intervention Recommender System Using Ml Algorithms

R. MD. Shafi, Y Sajani, M Saritha
Department of Artificial Intelligence & Data Science,
Aditya College of Engineering, Madanapalle, India.
E-mail: mdshafi.r@acem.ac.in

Abstract: Predicting student academic performance is essential for improving learning outcomes and identifying students who require timely support. This paper presents an adaptive student performance prediction and intervention recommender system using machine learning techniques. The proposed system analyzes academic records, attendance, and behavioral data to estimate student performance levels. Random Forest, XGBoost, ARD Regression, and Decision Tree algorithms are applied for classification and comparative evaluation. Experimental results demonstrate that the model achieves improved prediction accuracy and effectively identifies at-risk students at an early stage. Based on these predictions, the system suggests appropriate interventions to assist educators in providing targeted support. The proposed approach enables data-driven decision-making and contributes to enhancing overall student performance and academic success
***Keywords**: Student Performance Prediction, Machine Learning, Random Forest, XGBoost, Decision Tree, ARD Regression, Educational Data Mining, Recommender System*

I. Introduction

The use of data-driven approaches in education has significantly improved how student performance is monitored and evaluated. Today, educational institutions generate large amounts of data, including academic scores, attendance records, internal assessments, and behavioral information. With the growth of Educational Data Mining (EDM) and learning analytics, it has become possible to extract useful insights from this data and use them to support better academic decision-making.

Predicting student performance has become increasingly important due to rising concerns such as dropout rates, academic failures, and variations in student outcomes. Identifying students who are at risk at an early stage allows educators to provide timely support and improve retention. However, traditional methods that rely mainly on periodic exams and manual evaluation are often not sufficient to detect early signs of performance decline.

Machine Learning (ML) techniques provide an effective way to model complex relationships between different academic factors. In this work, a hybrid predictive framework is developed using Random Forest, XGBoost, Decision Tree, Ridge Regression, and Automatic Relevance Determination (ARD) Regression. Ensemble methods like Random Forest and XGBoost improve prediction accuracy, while Decision Trees offer better interpretability. Ridge and ARD regression help handle issues such as multi collinearity and feature relevance.

Although previous studies have made progress in this area, challenges such as overfitting, class imbalance, and limited integration of regression and classification methods still exist. Most existing approaches focus on either predicting scores or classifying performance levels, but not both together.

To address this gap, this study proposes an Adaptive Student Performance Prediction and Intervention Recommender System that combines both regression and classification approaches. The system predicts student performance and identifies risk levels while also suggesting suitable interventions. This approach aims to support early detection, improve academic outcomes, and enable more effective, data-driven decision-making in educational institutions.

II. LITERATURE SURVEY

The use of Machine Learning in education has made it easier to predict student performance with better accuracy. Earlier studies have shown that models such as Decision Trees, SVM, Random Forest, and XGBoost perform well for classification tasks, while deep learning approaches are effective when dealing with sequential or time-based data. In addition, ensemble methods have been widely used to improve overall performance by combining multiple models and enhancing generalization.

However, many existing approaches have certain limitations. Most of them focus only on classification tasks, such as predicting whether a student will pass or fail, without considering continuous score prediction. Common issues like overfitting, multicollinearity, and lack of interpretability are also observed. Moreover, only a few studies incorporate regularized regression techniques or provide meaningful intervention strategies based on predictions.

To overcome these limitations, this study proposes a hybrid framework that integrates both regression models (Ridge and ARD) and classification models (Decision Tree, Random Forest, and XGBoost). Regularization techniques are used to improve model stability and reduce overfitting, while ensemble methods help capture complex relationships in the data. The system also includes an intervention recommender to provide targeted academic support.

Overall, the proposed approach combines prediction and recommendation within a single framework, making it a more practical and reliable solution for analysing student performance.

A. AI-Based Early Prediction and Intervention for Student Academic Performance

Maram Khamis Al-Muharrami et al. [1] applied multiple machine learning algorithms, including SVM, Decision Tree, ANN, Naive Bayes, and k-NN, to predict student performance. Among these, SVM achieved the highest accuracy of 85.1% along with a strong F2-score. The study also introduced a desktop-based application that helps educators monitor student progress and provide timely academic support, making the system more practical for real-world use.

B. Student Performance Prediction Using Supervised Learning Techniques

Om Motvani et al. [2] presented a review of several studies that use supervised learning algorithms such as SVM, KNN, and Decision Trees. The paper compares different methodologies and evaluation metrics while also discussing common challenges like data imbalance and model selection. It highlights the increasing importance of machine learning in improving student performance prediction.

C. RNN-Attention Models for Early Prediction of Student Performance

Sukrit Leelaluk et al. [3] proposed a deep learning model based on RNN with Attention and Knowledge Distillation to predict student performance at early stages of a course. The model achieved improved recall and F1-score, demonstrating its effectiveness in identifying at-risk students early and supporting timely interventions.

D. Comparative Analysis of Machine Learning Algorithms for At-Risk Students

Zainab Mahmoud et al. [4] conducted a comparative study of various machine learning algorithms using academic and non-academic data. The results showed that Gaussian Process and Decision Tree models provided the best performance in terms of accuracy, R^2 score, and error minimization. The study emphasizes the importance of selecting suitable models for early warning systems.

E. Machine Learning-Based Engagement Prediction for Online Courses

Wanning Wang [5] analyzed student engagement prediction using Decision Tree, SVM, and Random Forest models. Among these, Random Forest performed best across multiple evaluation metrics. The study also included feature importance analysis to identify key factors affecting student engagement in online learning environments.

III. PROPOSED METHOD

The proposed system is designed to predict student academic performance and provide suitable intervention recommendations using a data-driven approach. It combines regression and classification techniques to improve prediction accuracy and support early identification of at-risk students.

A. Data Collection and Preprocessing

The dataset used in this study includes various academic and behavioral attributes such as study hours, attendance percentage, parental involvement, previous exam scores, sleep patterns, and participation in extracurricular activities. Each record represents an individual student and is used to predict their final academic performance.

Before model training, the dataset is preprocessed to ensure consistency and usability. Categorical variables are converted into numerical form using label encoding, and missing or inconsistent values are handled appropriately. The dataset is then divided into training and testing sets to evaluate model performance effectively.

B. System Implementation

The system is implemented using Python, which provides a flexible environment for data analysis and model development. Libraries such as Pandas and NumPy are used for data handling and numerical computations. Machine learning models are developed using Scikit-learn, while XGBoost is used for advanced gradient boosting. Visualization tools like Matplotlib are applied to better understand data patterns and relationships.

C. Model Development

The proposed framework integrates multiple machine learning algorithms to capture different aspects of student performance. Regression models such as Ridge Regression and ARD Regression are used to predict continuous academic scores, while Decision Tree, Random Forest, and XGBoost models are used for classification and pattern recognition.

Ensemble methods like Random Forest and XGBoost help improve prediction accuracy and reduce overfitting by combining multiple decision trees. At the same time, regularization techniques in Ridge and ARD regression help handle multicollinearity and improve model stability. This combination ensures a balance between accuracy, interpretability, and generalization.

D. Performance Evaluation

To evaluate the effectiveness of the models, both classification and regression metrics are used. A confusion matrix is applied to assess classification performance by comparing predicted and actual categories.

For regression analysis, metrics such as R^2 Score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are used. These metrics help measure how accurately the model predicts student scores. A model with higher R^2 and lower error values is considered more effective.

Table 1: A list of evaluation metrics used in this research

Metric	**Formula**
R^2 Score	$R^2 = 1 - (\Sigma(yi - \hat{y}i)^2 / \Sigma(yi - \bar{y})^2)$
Mean Squared Error (MSE)	$MSE = (1/n)\ \Sigma(yi - \hat{y}i)^2$
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{}((1/n)\ \Sigma(yi - \hat{y}i)^2)$
Mean Absolute Error (MAE)	$MAE = (1/n)\ \Sigma$

E. Intervention Recommendation

Based on the predicted performance, the system identifies students who are at risk and provides targeted intervention suggestions. These recommendations may include improving study habits, increasing attendance, or focusing on specific subjects. This component helps educators take timely actions to support student improvement.

F. Overall Framework

The proposed method integrates prediction and recommendation into a single system. By combining regression and classification models with intervention strategies, the system provides a more complete and practical solution for student performance analysis. It supports early detection, improves decision-making, and enhances academic outcomes.

IV. PROPOSED ALGORITHM

In the proposed system, student academic performance is predicted using a combination of regression-based machine learning models. The system processes student-related data such as academic records, attendance, and behavioral attributes to estimate performance outcomes. Initially, the data flows through a preprocessing stage where it is cleaned and transformed into a suitable format for model training. Similar to how routing protocols establish efficient paths, the system identifies the most optimal predictive model by evaluating multiple algorithms and selecting the best-performing one.

The algorithm follows a structured pipeline where data is first prepared and then passed through different regression models to learn patterns. Each model generates predictions, and based on performance metrics, the most accurate model is chosen for final prediction. The following steps describe the working of the proposed system:

1. The student dataset is collected, containing academic, attendance, and behavioral attributes.
2. The dataset is examined for missing values, inconsistencies, and irrelevant features, which are then handled appropriately.
3. Categorical attributes such as gender, degree, major, and hostel status are converted into numerical format using label encoding.
4. The dataset is divided into input features (independent variables) and the target variable (student performance score).

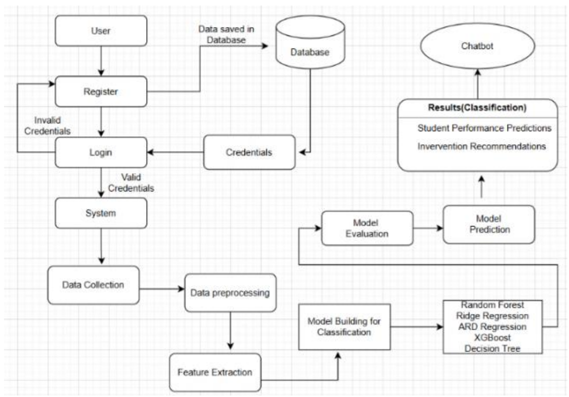

Figure 1: Data Preprocessing and Model Flow

5. The dataset is split into training and testing sets in an 80:20 ratio to ensure proper validation.
6. Multiple regression models, including ARD Regression, Ridge Regression, Decision Tree Regressor, Random Forest Regressor, and XGBoost Regressor, are trained using the training data.
7. Each model learns patterns between input features and student performance, capturing both linear and non-linear relationships.
8. The trained models are then evaluated using performance metrics such as R^2 score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and accuracy percentage.

Table 2: Comparison of R2, MSE, RMSE, MAE, Accuracy of Algorithms

.Algorithm	R^2 Score	MSE	RMSE	MAE	Accuracy (%)
ARD Regression	0.9091	76.2550	8.7324	6.9906	90.91%
Ridge Regression	0.9244	63.3781	7.9610	6.3858	92.44%
Decision Tree	0.8120	157.660 0	12.5563	9.8360	81.20%
Random Forest	0.9150	71.3058	8.4443	6.7561	91.50%
XGBoost	0.9139	72.1982	8.4969	6.7140	91.39%

9. A comparative analysis is performed, and the model with the highest R^2 score and lowest error values is selected as the optimal model.
10. The selected model is used to predict student academic performance on new or unseen data.
11. Based on the predicted results, insights are generated to help identify students who may require additional academic support.

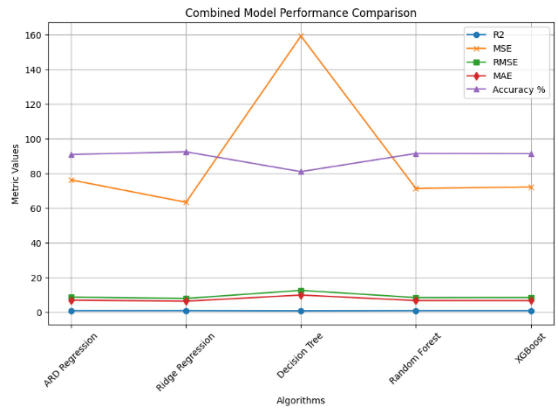

Figure 2: Combined Model Performance Comparision

Overall, the evaluation results consistently indicate that Ridge Regression outperforms other regression models across all performance metrics. The combination of high explanatory power and low prediction error confirms its effectiveness for student academic performance prediction.

V. SIMULATION RESULTS

The final stage of this project focuses on converting the selected machine learning model into a practical web-based prediction system. After comparing the performance of multiple regression algorithms, Ridge Regression was found to be the most effective model. This trained model was then integrated into a Python-based web application to allow real-time prediction of student performance.

To make the system efficient, the trained model is saved using model persistence techniques so that it does not need to be retrained every time the application runs. When the application starts, the saved model is loaded into the backend and remains ready to handle prediction requests.

The web application provides a simple and user-friendly interface where users can enter input data through a form. The inputs include important academic and personal factors such as study hours, attendance percentage, previous exam scores, sleep hours, parental involvement, access to resources, motivation level, and tutoring sessions. The input format is kept consistent with the features used during model training to ensure accurate and reliable predictions.

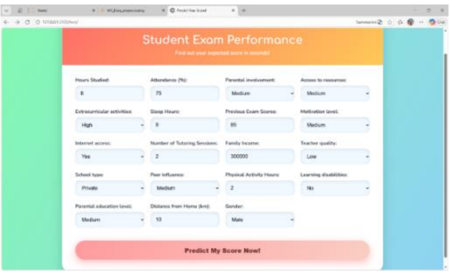

Figure 3: Input of Student Details

Once the user submits the form, the backend system validates the input values and performs necessary preprocessing steps, including encoding categorical variables and arranging features in the required order. The processed data is then passed to the trained Ridge Regression model for prediction. The model generates a predicted score representing the expected academic performance of the individual student.

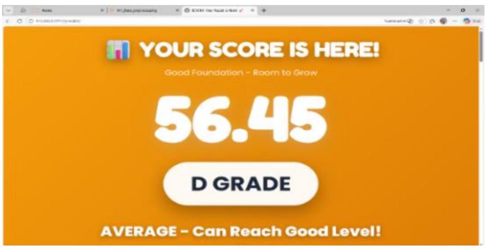

Figure 4: Score Prediction

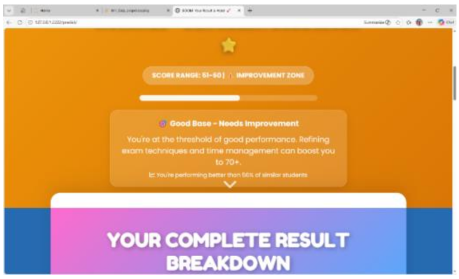

Figure 5: Student Performance

Figure 6: Optimal Daily Schedule

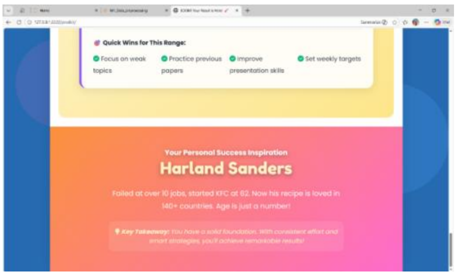

Figure 7: Example Inspiration for Student

The result is instantly displayed on the web interface, along with:

- Predicted Score (out of 100)
- Corresponding Grade (e.g., A, B, C, D)
- Performance Level (e.g., Excellent, Good, Average)
- Improvement Zone Indicator
- Personalized Action Plan

The system provides customized recommendations based on the predicted score range to guide students toward academic improvement.

Unlike bulk prediction systems, this application is specifically designed for individual-level prediction. It does not require CSV file uploads or batch processing. Instead, it focuses on delivering personalized insights and actionable feedback for a single student at a time, ensuring simplicity, clarity, and user-friendliness.

This seamless integration of the trained Ridge Regression model into a web application ensures efficient real-time prediction while maintaining accuracy, scalability, and practical usability.

Chatbot Integration

To improve user interaction and provide better academic support, a chatbot module was integrated into the web-based prediction system. The chatbot acts as an intelligent assistant that guides students based on their predicted performance. After the model generates the predicted score and performance level, the chatbot analyzes the results and

International Journal of Advanced Trends in Engineering, Science and Technology (IJATEST-ISSN:2456-1126) *Volume.11,Issue.2,2026*
DOI:10.22413/ijatest/2026/V11/i2/01

provides personalized suggestions. These include study strategies, time management tips, subject-specific improvement methods, and general motivation.

The chatbot works in real time and interacts directly with users through the web interface. Students can ask questions related to their academic performance, study plans, or exam preparation, and receive immediate responses. This makes the system more interactive and useful, transforming it from a simple prediction tool into a supportive learning assistant. Overall, the integration of the chatbot improves user engagement, accessibility, and provides a more adaptive learning experience.

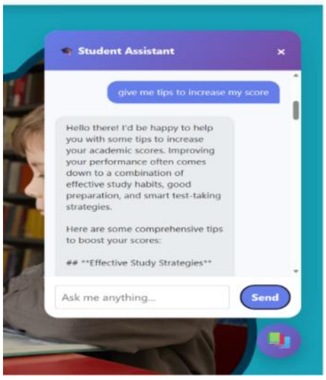

Figure 8: Chatbot Integration

VI. CONCLUSION

This study presents the development of a machine learning-based system for predicting student academic performance. Several regression algorithms, including ARD Regression, Ridge Regression, Decision Tree, Random Forest, and XGBoost, were implemented and evaluated to determine the most suitable model.

Among these, Ridge Regression achieved the best performance, with an R^2 score of 0.9244 and the lowest error values (MSE, RMSE, and MAE). This indicates that the dataset has largely linear characteristics, making Ridge Regression a stable and effective choice for prediction.

The selected model was successfully integrated into a web-based application that enables real-time prediction for individual students. In addition to predicting scores and performance levels, the system also provides personalized recommendations to help students improve their academic outcomes.

Overall, the proposed system bridges the gap between theoretical machine learning models and practical application. It serves as a useful decision-support tool for both students and educators, helping in performance monitoring and academic improvement.

VII. FUTURE WORK

In the future, the system can be further enhanced by incorporating advanced machine learning algorithms such as Support Vector Regression (SVR), Artificial Neural Networks (ANN), LightGBM, and CatBoost to improve prediction accuracy. Techniques like Isolation Forest can also be used to detect unusual patterns in student performance and identify sudden declines.

The chatbot module can be expanded to provide more intelligent and interactive support. By integrating Natural Language Processing (NLP) techniques, the chatbot can handle more complex queries, generate personalized study plans, and provide automated academic guidance. It can also analyze prediction results to offer adaptive learning suggestions, structured study schedules, and motivational support.

Additionally, the system can be extended to support real-time institutional data, bulk predictions for classrooms, and deployment as a mobile application. These improvements would make the system more scalable and transform it into a comprehensive academic analytics and student support platform.

REFERENCES

[1] Khan, I., et al., *An artificial intelligence approach to monitor student performance and devise preventive measures.* Smart Learning Environments, 2021. 8(1): p. 1–18.

[2] Zeineddine, H., U. Braendle, and A. Farah, *Enhancing prediction of student success: Automated Machine Learning approach.* Computers & Electrical Engineering, 2021. 89: p. 106903.

[3] Hashim, A.S., W.A. Awadh, and A.K. Hamoud, *Student performance predictions a model based on supervised machine learning algorithms.* IOP Conference Series: Materials Science and Engineering, 2020.

[4] Lau, E.T., L. Sun, and Q. Yang, *Modelling, prediction and classification of student academic performance using artificial neural networks.* SN Applied Sciences, 2019. 1(9): p. 982.

[5] Namoun, A. and A. Alshanqiti, *Predicting student performance using data mining and learning analytics techniques: A systematic literature review.* Applied Sciences, 2020. 11(1): p. 237.

[6] Yağcı, M., *Educational data mining: prediction of students' academic performance using machine learning algorithms.* Smart Learning Environments, 2022. 9(1): p. 11.

[7] Sukhbaatar, O., T. Usagawa, and L. Choimaa, *An artificial neural network based early prediction of failure-prone students in blended learning course.* International Journal of Emerging Technologies in Learning (iJET), 2019. 14(19): p. 77–92.

[8] Alhazmi, E. and A. Sheneamer, *Early predicting of students performance in higher education.* IEEE Access, 2023. 11: p. 27579–27589.

[9] Musso, M.F., C.F.R. Hernández, and E.C. Cascallar, *Predicting key educational outcomes in academic trajectories: a machine-learning approach.* Higher Education, 2020. 80(5): p. 875–89.