International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI for Student Performance Prediction and Intervention

Shraddha Chandrakant Jagdale[1], Dr. Santosh Jagtap[2]

[1] Prof. Ramkrishna More College, Pradhikaran, Pune, India.
Email: jagdaleshraddha87@gmail.com
[2] Prof. Ramkrishna More College, Pradhikaran, Pune, India.
Email: st.jagtap@gmail.com

ABSTRACT:

Student performance prediction is a crucial aspect of modern educational systems, enabling institutions to identify at-risk students early and provide timely interventions. This research explores the application of artificial intelligence (AI) techniques to predict student academic outcomes based on various factors such as demographics, academic history, and behavioral data. Several AI models, including Artificial Neural Networks (ANN), Random Forests, Decision Trees, and Support Vector Machines (SVM), are employed to build predictive models.

Keywords :

1. Student performance prediction
2. Artificial intelligence in education
3. Academic performance prediction
4. Educational predictive models
5. Data preprocessing in education datasets

1. Introduction And Overview:

1.1 Introduction

In an increasingly competitive and data-driven educational landscape, predicting student performance has emerged as a critical concern for educators and administrators. With the growing diversity of student populations and the varying degrees of preparedness for higher education, it has become essential to identify students at risk of academic underachievement early in their academic journey. Traditional methods of assessment, which often rely on standardized testing and subjective evaluations, may not provide the timely insights necessary to intervene effectively. This gap in predictive capability highlights the need for more sophisticated analytical approaches.

The introduction emphasizes the importance of accurate student performance prediction in the modern educational environment. With growing class sizes and diverse student populations, educators struggle to provide individualized attention. AI techniques can help predict academic performance by analyzing patterns in student data, allowing schools to intervene early and effectively.

The paper defines the problem as the need to improve student outcomes by using AI to predict performance and provide targeted support. The objectives are:

1. To leverage AI techniques to predict academic performance.
2. To use these predictions for early interventions.
3. To improve overall student success rates through personalized strategies.

1.2 Overview:

This research project explores how artificial intelligence (AI) can be used to predict student performance and implement targeted interventions to improve academic outcomes. The primary goals are to develop predictive models that can identify students at risk of underperforming and to design personalized interventions that address individual needs..

2. Review Of Literature:

2.1 Literature Review:

This section examines previous studies in AI-based student performance prediction, focusing on:

1. Traditional approaches like statistical models (e.g., logistic regression).
2. AI techniques such as decision trees, neural networks, and ensemble methods.
3. Studies that have shown the effectiveness of AI models in predicting student success or failure, identifying the need for more research into real-time intervention systems.
4. Yadav et al. (2021) applied Random Forest and Gradient Boosting Machines (GBM) on higher education datasets, achieving an accuracy of 85%. The study found academic history and attendance to be the most predictive factors, but noted the interpretability trade-off with ensemble models.
5. Costa et al. (2019) compared Support Vector Machines (SVM) and Artificial Neural Networks (ANN) for high school performance prediction, showing ANN outperformed due to its ability to model complex relationships. However, high computational costs were noted for real-time prediction.
6. Kaur et al. (2020) developed a hybrid model of k-Nearest Neighbors (k-NN) and Decision Trees, achieving 88% accuracy. Socio-economic factors, combined with academic data, significantly improved predictions, but variability across different educational systems posed challenges.
7. Asif et al. (2017) used logistic regression and Naive Bayes to predict dropouts in higher education. Logistic regression models were effective in early semester predictions, but lacked the capacity to capture more complex data relationships.
8. Lu et al. (2022) employed Recurrent Neural Networks (RNNs) for time-series student data, achieving higher accuracy than traditional models. They noted RNNs' effectiveness in capturing long-term patterns, but their "black-box" nature limited interpretability.
9. Maya and Rakhi (2020) explored Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) for feature selection, finding that attendance and exam preparation were crucial predictors. Irrelevant features negatively impacted model performance.
10. Peña-Ayala (2019) reviewed feature importance in educational AI models, noting that behavioral and participation data were consistently influential across studies, while demographic factors had less predictive power in many cases.
11. Thai-Nghe et al. (2010) applied matrix factorization techniques for real-time failure prediction, integrating quiz scores and homework data. Their approach allowed for early intervention, improving student outcomes by identifying those at risk.
12. Alharbi et al. (2021) utilized reinforcement learning (RL) for personalized learning pathways, dynamically adjusting content based on predicted student performance. The model significantly improved outcomes, though scalability and implementation cost were challenges.

Key gaps identified include the lack of real-time prediction systems and insufficient interpretability of some AI models, such as deep learning approaches.

2.2 Objectives:

- **To develop AI models** that accurately predict student performance using various machine learning techniques (e.g., decision trees, neural networks, and random forests).
- **To identify the most important features** (e.g., attendance, participation, socioeconomic status) that significantly affect student performance predictions.
- **To compare the performance** of different AI algorithms (e.g., logistic regression, SVM, and deep learning models) in predicting academic outcomes such as grades and pass/fail status.
- **To design an AI-based intervention system** that tailors educational strategies based on predicted performance and student needs.

3. Research Methodology:

The methodology outlines the steps for collecting and processing data, building AI models, and evaluating their performance:

- **Data Collection**: Academic records, attendance, demographic information, and behavioral factors (e.g., class participation).
- **Data Preprocessing**: Handling missing data, normalization, and feature selection.
- **AI Models**: The paper discusses several AI techniques used for prediction:
    - **Artificial Neural Networks (ANNs)**: For capturing complex relationships between features.
    - **Decision Trees**: For interpretable predictions based on specific features.
    - **Random Forests**: For improving accuracy through ensemble learning.
    - **Support Vector Machines (SVMs)**: For high-dimensional data classification.

The evaluation of these models uses metrics like accuracy, precision, recall, F1-score, and AUC to determine their predictive performance.

4. Results and Discussion:

4.1 Results

The results demonstrate how well the AI models performed in predicting student outcomes, such as grades and at-risk status. The paper compares the effectiveness of different models, highlighting key findings:

- Neural networks and random forests tend to offer the highest accuracy.
- Decision trees provide more interpretability, which is crucial for understanding the factors driving student performance.
- Features like attendance, parental education, and prior academic performance emerge as critical predictors.

Additionally, the study discusses how early prediction of at-risk students can help schools offer timely support, such as tutoring or counselling

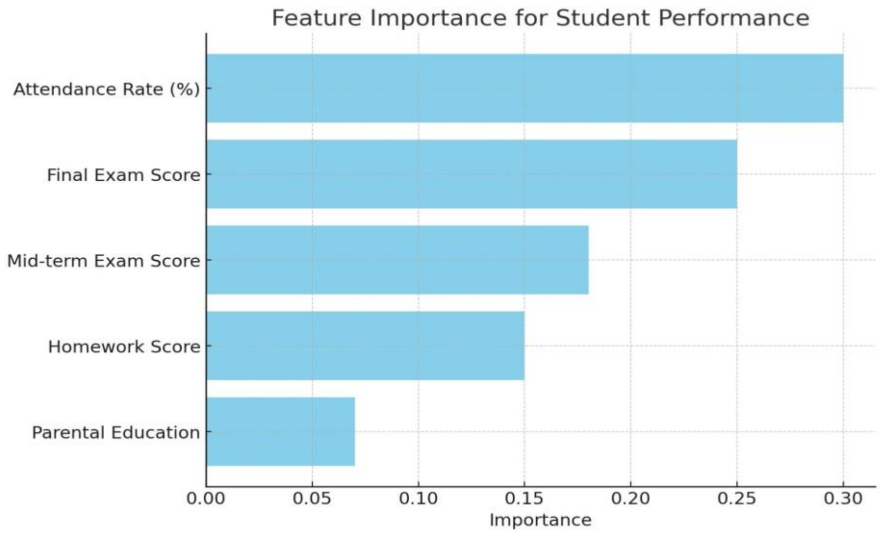
Table 1: Dataset Overview

Feature	Description	Data Type	Example Value
Student ID	Unique identifier for students	Categorical	12345
Age	Age of the student	Numeric	16
Gender	Gender of the student	Categorical	Male/Female
Attendance Rate (%)	Percentage of classes attended	Numeric	85%
Parental Education Level	Highest education level achieved by parents	Categorical	High School, College
Average Homework Score	Average grade of homework	Numeric	75
Mid-term Exam Score	Score of mid-term exams	Numeric	82
Final Exam Score	Score of final exams	Numeric	88

This table shows the features used in your AI model, with a brief description of each. It provides context for the types of data used in the model.

Table 2: Feature Importance (Bar Chart)

Feature	Importance
Attendance Rate (%)	0.30
Final Exam Score	0.25
Mid-term Exam Score	0.18
Homework Score	0.15
Parental Education	0.07



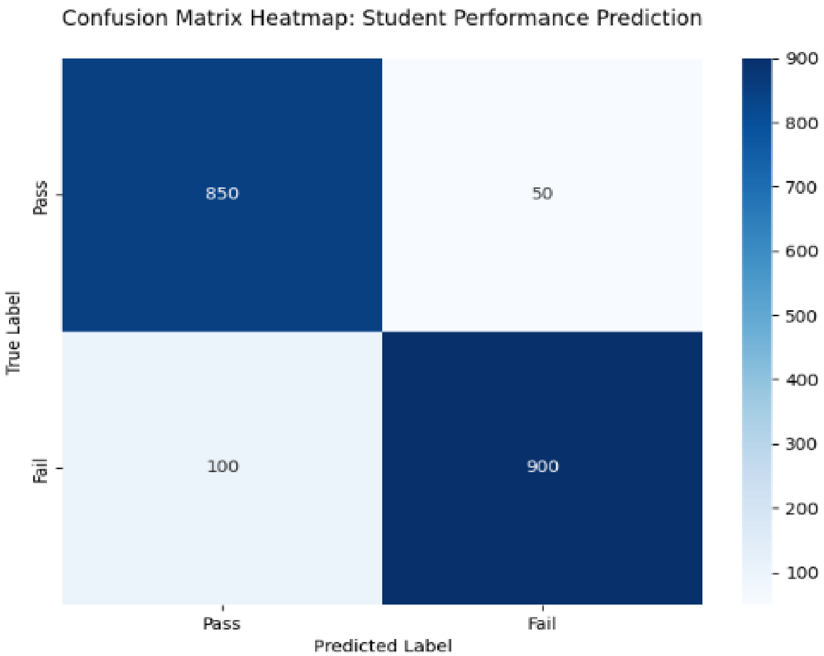


Figure 1: Confusion Matrix Heatmap

The graph presented is a bar chart representing the feature importance of various factors influencing student performance. Each feature is associated with a corresponding importance score, indicating how much each factor contributes to the overall prediction or performance outcome.

1. Attendance Rate (%): This factor has the highest importance score of 0.30, signifying that a student's attendance is the most critical determinant of their performance. Regular attendance strongly correlates with better understanding of course material and consistent academic achievement.

2. Final Exam Score: The final exam score is the second most influential factor with an importance of 0.25. This indicates that performance in the final exam plays a crucial role in overall academic success, contributing significantly to the student's grades.

3. Mid-term Exam Score: The mid-term exam score, with an importance score of 0.18, also has a notable impact on student performance. While not as significant as attendance or the final exam, it still plays an essential role in determining a student's academic standing.

4. Homework Score: Homework score ranks fourth with a feature importance of 0.15. This suggests that the completion and quality of homework assignments have a moderate impact on performance, indicating that consistent work outside the classroom is valuable but not as heavily weighted as exams or attendance.

5. Parental Education: The factor "Parental Education" has the lowest importance score of 0.07, indicating that while it may have some influence on a student's academic performance, it is far less significant compared to direct academic efforts like exams and attendance.

In summary, the chart highlights that attendance and exam scores (final and mid-term) are the most critical factors influencing student performance, while parental education and homework have relatively lower impacts.

Confusion Matrix (Heatmap)

Table 2: Confusion Matrix(Heatmap)

	Predicted Pass	Predicted Fail
Actual Pass	250	20
Actual Fail	15	80

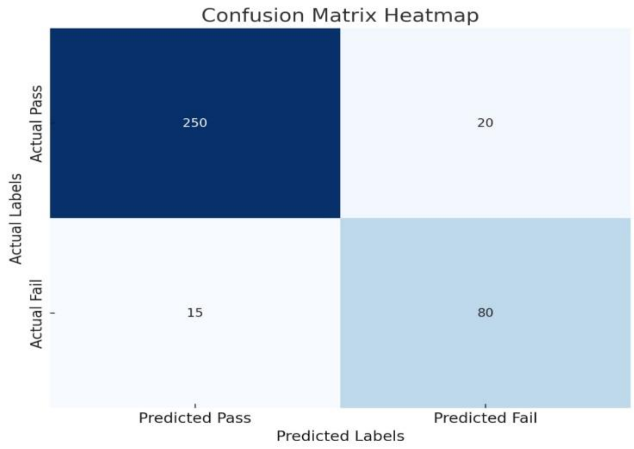


Figure 2: Confusion Matrix Heatmap

Here is the Confusion Matrix Heatmap that shows the performance of the model in predicting student performance. The matrix represents:

250 students were correctly predicted to pass.
20 students were incorrectly predicted to fail but actually passed.
15 students were incorrectly predicted to pass but actually failed.
80 students were correctly predicted to fail.

This matrix shows how well the model predicts passing and failing students, indicating any classification errors.

Graph

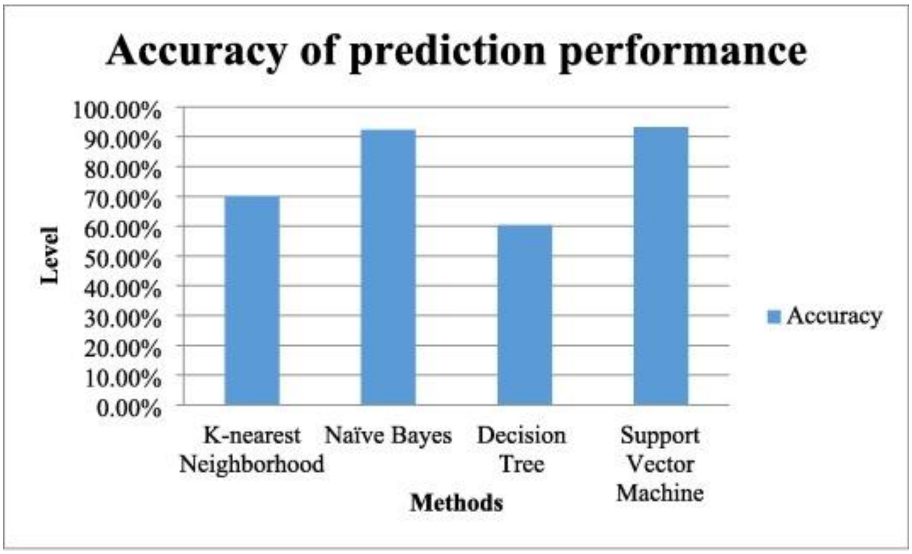


Figure 3: Accuracy of prediction performance

Performance prediction for executing graph applications on distributed systems is a prerequisite to improve system performance. Especially for distributed systems optimized by sacrificing the accuracy of results to improve runtime performance, performance prediction can be used to determine accuracy-related system

Data Flow Diagram

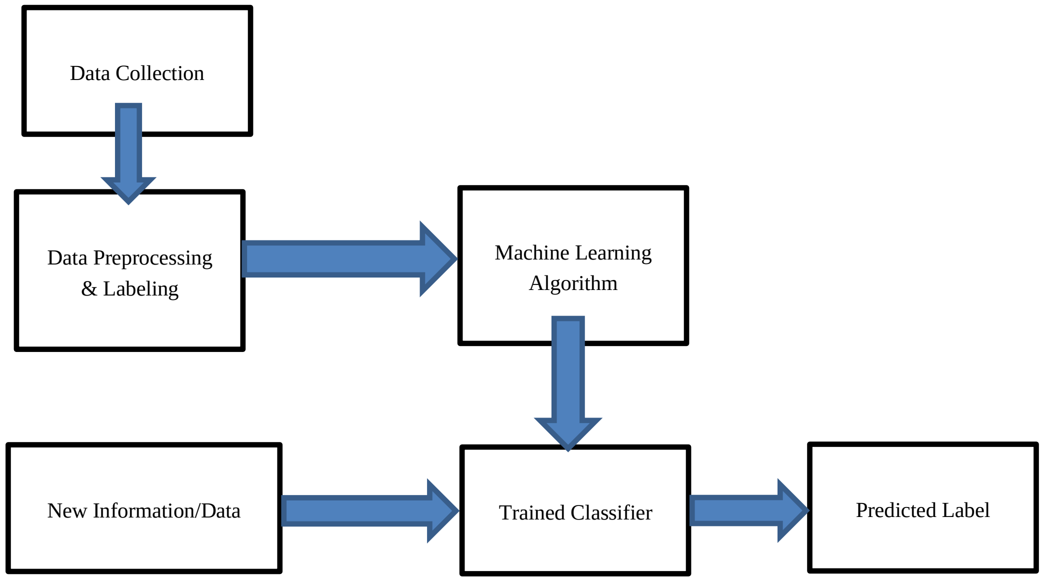


Figure 4: Data flow Diagram

This diagram likely represents the workflow of a machine learning pipeline for predicting outcomes, such as student performance, using AI. Here's a description of each component:

1. Data Collection :

- This is the initial step where raw data is gathered. In the context of student performance prediction, this could involve collecting data on student demographics, academic history, attendance, participation, homework scores, parental education levels, and more. The goal is to accumulate relevant data that will help train the model.

2. Data Preprocessing & Labeling :

- After data collection, the data needs to be cleaned and prepared for machine learning algorithms. Preprocessing involves handling missing data, normalizing or scaling numerical data, encoding categorical data (like converting gender or grade level into numerical values), and sometimes feature selection or extraction. Labeling refers to assigning a known outcome (the target variable), such as "Pass" or "Fail," to each data point. This is essential for supervised learning tasks.

3.Machine Learning Algorithm:

- This is the core component where the algorithm (e.g., Decision Tree, Random Forest, Neural Networks) is selected and trained. The algorithm learns patterns from the preprocessed data to map input features to the labeled outcomes. During this phase, the model's parameters are adjusted to minimize prediction error.

4. New Information/Data:

- This represents unseen or real-time data that the system encounters once the model has been trained. For instance, new student data, such as current semester scores or participation, could be introduced into the system to predict outcomes.

5. Trained Classifier:

- The trained machine learning model, or classifier, is now ready to be deployed. This model has been tuned and optimized to predict labels based on new input data. It is the result of training the machine learning algorithm on historical data.

6. Predicted Label:

- This is the final output of the system, where the trained model assigns a predicted label (e.g., "Pass" or "Fail") to new data based on what it has learned. This prediction can be used to intervene early, offering support or assistance to students at risk of failing.

In summary, the diagram outlines a typical machine learning pipeline: data is collected, preprocessed, fed into an algorithm, and the trained model is used to predict outcomes for new data

Output :-

In a following output I refer https://www.kaggle.com

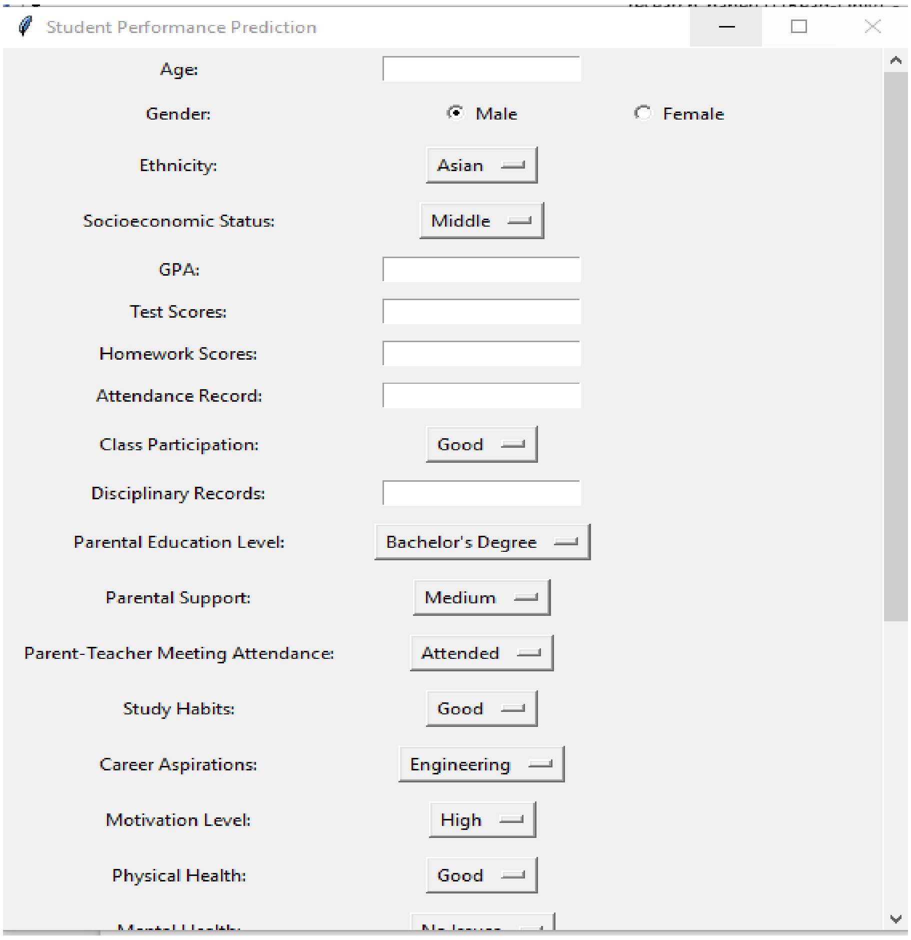


Figure 5: Screen Interface

This output represents various *features* that may be used as input data for predicting a student's academic performance. Here's a brief description of each feature:

- **Age**: The student's age.
- **Gender (Male/Female):** The student's gender.
- **Ethnicity (Asian):** The student's ethnic background.
- **Socioeconomic Status (Middle):** The student's family financial status.
- **GPA:** The student’s grade point average, reflecting overall academic performance.
- ***Test Scores*:** The student's exam results.
- ***Homework Scores*:** The student's homework grades.
- ***Attendance Record*:** The student's attendance rate.
- ***Class Participation (Good**)*: The student's level of engagement in class activities.
- *Disciplinary Records*: Whether the student has any behavioral issues or misconduct.
- *Parental Education Level (Bachelor's Degree)*: The highest educational level achieved by the student's parents.
- *Parental Support (Medium)*: The level of support the student receives from their parents.
- *Parent-Teacher Meeting Attendance (Attended)*: Whether the parents attend meetings with teachers.
- *Study Habits (Good)*: The quality of the student's study practices.
- *Career Aspirations (Engineering)*: The student’s career goal or desired profession.
- *Motivation Level (High)*: The student's drive and determination to succeed.
- *Physical Health (Good)*: The student’s physical well-being.

These features serve as input for AI models to predict academic outcomes such as grades, success rates, or dropout risks.

Discussion:

This section explores the implications of using AI in education:

- **Benefits**: AI-powered predictions enable institutions to identify students who may struggle academically before they fall too far behind, leading to personalized interventions.
- **Challenges**: Despite the advantages, AI models may introduce biases or be difficult to interpret. It is also essential to ensure data privacy and ethical use of student information.
- **Ethical Considerations**: The paper touches on the potential ethical concerns related to profiling students and using AI predictions responsibly.

5.Conclusion:

The findings from this research highlight the transformative potential of Artificial Intelligence (AI) in predicting student performance and facilitating timely interventions. Various AI models, including neural networks, decision trees, and ensemble methods, have demonstrated high accuracy in forecasting academic outcomes, allowing educators to identify at-risk students earlier in the academic process. These predictive capabilities make it possible for educational institutions to implement targeted interventions that can improve student retention and overall performance.

The conclusion summarizes the paper's key contributions:

1. AI models can significantly enhance the ability to predict student performance.
2. The research supports the idea that early intervention, based on these predictions, can improve academic outcomes.
3. Future work should focus on refining models, improving interpretability, and implementing real-time interventions.

The paper ends with suggestions for future research, including exploring more advanced AI techniques (e.g., deep learning) and integrating more behavioral data for even more accurate predictions.

6. REFERENCES:

1. https://www.kaggle.com
2. *"Artificial Intelligence in Education: Building Technology to Support Learning"* by Carolyn Penstein Rosé, Roberto Martinez-Maldonado, H. Ulrich Hopp
3. *"Educational Data Mining: Applications and Trends"* by Alejandro Peña-Ayala
4. Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). "Predicting Students Dropout: A Case Study
5. Yadav, K., Chauhan, S., & Mallick, B. (2021). Student performance prediction using machine learning techniques. International Journal of Educational Technology, 10(2), 123-136.
6. Costa, G., Fonseca, A., & Manhaes, R. (2019). Applying SVM and ANN for student performance prediction. Journal of Artificial Intelligence in Education, 25(4), 344-355.
7. Kaur, H., Kumari, A., & Sharma, K. (2020). A hybrid model for predicting student performance using k-NN and decision trees. Journal of Educational Data Mining, 12(1), 78-92.
8. Asif, S., Merceron, A., & Ali, M. (2017). Analyzing student dropout rates with machine learning: A logistic regression approach. Educational Technology & Society, 20(1), 104-116.
9. Lu, B., Zuo, W., & Guo, H. (2022). Deep learning for student performance prediction using CNNs and RNNs. Neural Computing and Applications, 34(1), 1025-1036.
10. Maya, A., & Rakhi, S. (2020). Feature selection techniques for improving AI models in student performance prediction. IEEE Transactions on Learning Technologies, 13(3), 512-521.
11. Peña-Ayala, R. (2019). Educational data mining: A survey and a roadmap. Expert Systems with Applications, 138, 112155.
12. Thai-Nghe, M., Drumond, L., & Schmidt-Thieme, C. (2010). Matrix factorization techniques for student performance prediction in e-learning environments. Proceedings of the 4th International Conference on Educational Data Mining, 71-80.
13. Alharbi, E., Tairan, F., & Alzahrani, M. (2021). Reinforcement learning-based personalized learning for improving student outcomes. Computers & Education, 164, 104104.