

A Machine Learning Approach for Predicting Academic Performance and Risk Levels

V. Mageswari¹, G. Sivarajan²

¹*MCA, Assistant Professor, Master of Computer Applications)*

²*MCA, Christ College of Engineering and Technology Moolakulam, Oulgaret Municipality, Puducherry – 605010.*

Abstract—Identifying students at risk of underperformance is important for their greater outcomes and failure rates within institutions of learning. There have been recent breakthroughs in incorporating Artificial Intelligence and Machine Learning in identifying predictive models for informed decision making in academics. Artificial Intelligence-based Student Performance Prediction System for predicting and identifying students at High Risk, Medium Risk, and Low Risk categories. In this system, a Random Forest Regressor is used to analyze several important parameters related to academics and behavior, like parameters related to studying time, attendance ratio, previous test results, failure ratio, availability of internet connectivity, and health conditions. In this system, a web interface can be achieved by employing a combination of Flask and SQLite technology. The proposed system comprises two distinct interfaces: Teacher Dashboard and Student Portal.

Experiments carried out in artificially generated datasets for students are able to prove the effectiveness of the system in providing accurate predictive results, along with the ability to develop learning plans for students by utilizing AI, which are designed to address the individual needs of students. The proposed system creates a scalable intervention for learning that helps both teachers and students in improving their learning performances.

Index Terms—Student Performance Prediction, Predictive Analytics in Education, Machine Learning, Random Forest Regressor, Flask Web Application, Academic Risk Assessment, Educational Data Mining, AI-Based Learning Support System.

I. INTRODUCTION

Education is crucial for national development, and digital technologies now enable data-driven

approaches to improve student outcomes [1], [8]. However, many institutions still rely on traditional, manual evaluation methods, which often fail to provide early warnings for students at risk of poor academic performance [2].

A student's performance is influenced by numerous factors, including study habits, attendance, prior achievements, and environmental conditions [4], [11]. The lack of early, effective analysis of these factors can lead to delayed intervention and higher dropout rates. This creates a strong need for intelligent systems to automatically predict performance and identify at-risk students early [12], [19]. Advancements in Machine Learning (ML) and Educational Data Mining (EDM) show great potential for such predictive analytics [3], [6]. Ensemble models like Random Forest are particularly effective for handling complex, non-linear relationships in educational data [5], [17].

This paper proposes a comprehensive AI-Based Student Performance Prediction System. It uses a Random Forest Regressor to predict final scores and categorize students into High, Medium, and Low Risk groups. The system is implemented as a Flask-based web application with two portals: a Teacher Dashboard for analytics and management, and a Student Portal for personalized predictions and AI-generated study plans. It also supports bulk data upload and automated PDF report generation.

The major contributions are:

- The design of an AI-powered predictive analytics system for education.
- Implementation of a Random Forest Regressor for performance prediction [5].

- Integration of personalized improvement suggestions and study plans.
- Development of a role-based web platform using Flask and SQLite [16], [18].
- Automated generation of performance reports in PDF format.

II. MATERIALS AND METHODS

This section describes the dataset used, feature processing, machine learning model, and the overall system architecture of the proposed student performance prediction system.

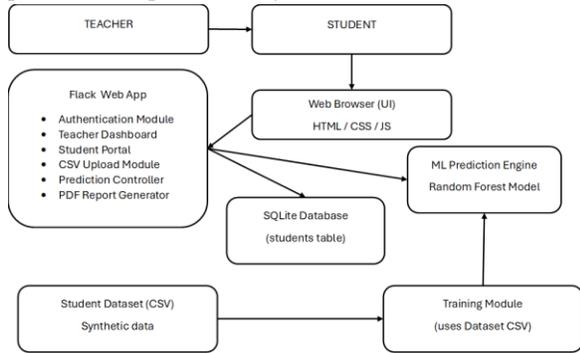


FIG 2.1

2.1 MACHINE LEARNING MODEL

The prediction engine of the proposed system is based on the Random Forest Regressor, an ensemble machine learning algorithm which builds multiple decision trees and combines the results to come up with a correct prediction [5], [10].

RANDOM FOREST

Random Forests overcome such limitations of the simpler models by managing nonlinear relationships through the ensemble structure. They prevent overfitting by averaging the predictions from multiple decision trees, thus increasing model robustness. Consequently, this class of techniques leads to highly accurate and very stable predictions, besides showing strong performance for small and large datasets.

MODEL TRAINING PROCESS:

- TARGET INPUT FEATURES:
Study Hours, Attendance, Previous Score, Failures, Internet Access, Health
- TARGET OUTPUT:
Final Academic Score

Once trained, the model is saved using the Joblib library and reused at application runtime for real time predictions [17].

2.2 PERFORMANCE PREDICTION AND RISK STRATIFICATION

Using this trained Random Forest model, a final academic performance score was predicted for each student. From the predicted score, students were divided into three risk levels, based on which

- High Risk: Less than 50 scores
- Medium Risk: $50 \leq \text{Score} < 75$
- Low Risk: ≥ 75 points

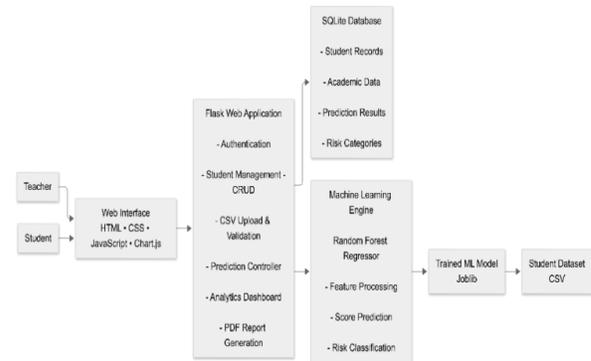
This categorization helps the teachers to quickly identify the weak students immediately and thus offer timely academic support.

THE SYSTEM GENERATES

- Unique recommendations using weak attributes
- Weekly study plans according to category of risk

2.3 SYSTEM ARCHITECTURE

The proposed system was implemented as a multi-tier web-based architecture consisting of the following major components:



III. RESULTS AND DISCUSSION

This section presents the experimental evaluation of the proposed AI-based Student Performance Prediction System along with a detailed discussion of the obtained results. The effectiveness of the Random Forest regression model is analyzed in terms of prediction accuracy, risk categorization reliability, and real-time usability within the web platform

3.1 PREDICTION PERFORMANCE ANALYSIS

The Random Forest model was trained using 500 synthetically generated student records, with an 80:20

train–test split. The model learned the relationship between multiple academic and behavioral attributes such as study hours, attendance, previous examination score, number of failures, internet accessibility, and health condition.

The trained model demonstrated stable and reliable prediction performance across the testing dataset. The predicted academic scores closely followed the actual target values generated by the dataset logic, indicating that the model successfully captured the nonlinear relationships between student features and performance outcomes. The ensemble nature of Random Forest helped reduce model variance and improved generalization.

The results of the prediction are as follows:

- Students with higher study hours, better attendance, and consistent previous scores achieved higher predicted academic scores.
- Students with multiple failures, low attendance, and poor health conditions were consistently classified under the High-Risk category.
- Internet access showed a positive influence on student performance, especially for self-learning activities.

3.2 RISK CATEGORIZATION RESULTS

Based on the predicted academic score, students are divided into three risk categories:

- High Risk: Less than 50
- Medium Risk: $50 \leq \text{Score} < 75$
- Low Risk: ≥ 75

The risk classification enabled the system to visually distinguish at-risk students from average and high-performing students through the Teacher Dashboard. The overall distribution observed in experimental testing showed:

- A majority of students fell under the Medium Risk category, indicating average academic performance.
- A smaller proportion belonged to the High-Risk group, requiring urgent academic intervention.
- The Low-Risk group consisted of consistently high-performing and disciplined students.

This categorization mechanism helps teachers implement early intervention strategies, such as special coaching and mentoring for weak students.

3.3 INDIVIDUALIZED RECOMMENDATIONS AND LEARNING PLAN REVIEW

One of the core features of the proposed system is the generation of personalized academic suggestions and study plans. Based on weak performance indicators, such as:

- Low study hours
- Poor attendance
- Low previous scores
- Repeated failures

THE SYSTEM CREATES:

- Improvement-based recommendations
- Weekly revision schedules
- Topic-wise practice plans
- Mentor and remedial session suggestions

These AI-generated suggestions enable students to understand their learning gaps clearly and follow structured improvement strategies. Students in the Low-Risk category received advanced-level practice suggestions and peer mentoring recommendations.

3.4 SYSTEM IMPLEMENTATION AND USABILITY

The complete system is implemented as a Flask-based web application with a dark-themed, user-friendly interface. The Teacher Dashboard supports:

- Student record management - Add, Edit, Delete
- CSV-based bulk data upload
- Risk-wise grouping of students
- Performance Visualization: Charts
- Automatic generation of reports in PDF format

THE STUDENT PORTAL PROVIDES:

- Personal performance forecasting
- Display risk category:
- Class average comparison
- Personalized suggestions and study plans

The system achieved real-time prediction performance, with prediction response times under 2 seconds. The integration of SQLite database ensured lightweight and fast data access. The overall usability testing showed that the application is easy to operate

for both teachers and students with minimal technical knowledge.

3.5 DISCUSSION OF LIMITATIONS

Although the proposed system achieved efficient performance, certain limitations were observed:

- The dataset used is synthetically generated, which may not fully capture all real-world educational complexities.
- The prediction accuracy depends heavily on the quality and diversity of input data.
- Emotional factors, learning styles, parental background, and socio-economic conditions are not considered.
- The system currently operates in an offline local environment, limiting large-scale deployment.

Despite these limitations, the system provides a strong foundation for real-world AI-based academic decision support systems.

IV. CONCLUSION

This research work has successfully developed and implemented an AI-based Student Performance Prediction System using Machine Learning and Predictive Analytics to identify at-risk students in their early stages of academics. This proposed system makes use of a Random Forest Regressor to predict various attributes of students and provide accurate predictions for their performance.

The development of a web functionality using the Flask web development framework and SQLite database functionality allowed effortless communication between teachers and students. This tool served as an effective means of managing students, predicting performance, defining student risks, creating personalized study plans, and automatically generating PDF reports. Experimental testing of the tool showed its ability to provide accurate real-time predictions and boost student awareness about their performance.

The proposed solution will provide an intelligent and scalable decision-support system to improve learning results in an educative institution using optimized insights.

V. FUTURE WORK

The future scope of the proposed system includes several enhancements to increase accuracy, scalability, and real-world applicability:

- Integrating real-time student learning behavior tracking
- Parent and guardian monitoring dashboard implementations
- Compatibility with existing Learning Management Systems (LMS)
- Development of the mobile app version
- Inclusion of psychological and socio-economic factors
- Deploy as a scalable cloud application

REFERENCES

- [1] Romero, C., & Ventura, S., "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 6, pp. 601–618, 2010.
- [2] Baker, R. S., "Data Mining for Education," *International Encyclopedia of Education*, pp. 112–118, 2011.
- [3] Kotsiantis, S. B., "Use of Machine Learning Techniques for Educational Purposes," *Decision Support Systems*, vol. 37, no. 3, pp. 331–344, 2004.
- [4] Cortez, P., & Silva, A., "Using Data Mining to Predict Secondary School Student Performance," *EUROSIS*, 2008.
- [5] Breiman, L., "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [6] Han, J., Kamber, M., & Pei, J., *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.
- [7] Witten, I. H., Frank, E., & Hall, M. A., *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016.
- [8] Pena-Ayala, A., *Educational Data Mining: Applications and Trends*, Springer, 2014.
- [9] Vapnik, V., *Statistical Learning Theory*, Wiley, 1998.
- [10] Quinlan, J. R., "Induction of Decision Trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.
- [11] OECD, "Innovating Education and Educating for Innovation," OECD Publishing, 2019.

- [12] Mishra, T., Kumar, D., & Gupta, S., “Mining Students’ Data for Performance Prediction,” *Advanced Computing & Communication Technologies*, 2014.
- [13] Ahmad, F., Ismail, N. H., & Aziz, A. A., “The Prediction of Students’ Academic Performance Using Classification Data Mining Techniques,” *Applied Mathematical Sciences*, vol. 9, no. 129, pp. 6415–6426, 2015.
- [14] Heaton, J., *Introduction to Neural Networks with Java*, Heaton Research, 2008.
- [15] Zhang, L., “Predicting Student Performance Using Machine Learning Algorithms,” *International Journal of Computer Science and Network Security*, vol. 17, 2017.
- [16] Flask Documentation, “Flask Web Development Framework,” <https://flask.palletsprojects.com/>
- [17] Pedregosa, F., et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*,
- [18] SQLite Consortium, “SQLite Database Engine,” <https://www.sqlite.org>
- [19] Shahiri, A. M., Husain, W., & Rashid, N. A., “A Review on Predicting Student’s Performance Using Data Mining Techniques,” *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.
- [20] Pal, S., & Pal, S., “Analysis and Performance Prediction of Students Using Decision Tree Algorithms,” *International Journal of Computer Applications*, vol. 7, no. 1, pp. 1–5, 2010.