

EduRisk MIS: An AI-Driven Management Information System for Learning Equity and Student Dropout Risk Prediction Using Weighted Multi-Factor Analysis

Sharvil Ganesh Nikam¹, Siddhi Ganesh Salunke², Shravani Yogesh Pardeshi³,
Dr. Pritesh Patil⁴, Dr. Mrunal Pathak⁵

UG Students, Department of Information Technology¹⁻³

Associate Professor, Department of Information Technology⁴

Assistant Professor, Department of Information Technology⁵

AISSMS Institute of Information Technology, Pune, India

sharvilgnikam15@gmail.com

Abstract: Education in developing countries is facing a big problem with students dropping out a lot, and its all tied to those socioeconomic differences that make it harder for kids from poorer backgrounds. I think the main issue is how these inequalities keep showing up in schools, affecting who gets a fair shot at learning. This paper talks about a new system called EduRisk MIS, which is basically an AI tool to spot students who might fail or leave school early. It helps with coming up with ways to step in and fix things, and also points out the gaps in equity based on family money situations. The way it works is by pulling together stuff like how often kids show up, their grades so far, and how engaged they seem, then giving each one a risk score thats weighted somehow. There are these analytical parts built in, like a heatmap that shows attendance over twelve months, which could be useful for seeing patterns. Then theres a tool for breaking down grades by level, and a leaderboard to compare how students are doing overall. It has a CSV importer to connect with other school software easily, and you can export reports for single students. Oh, and dark mode for when its late at night or something. Access is controlled by roles, so administrators see everything, teachers get their class views, NGO people can analyze broader trends, and students might have a simple dashboard. That setup makes sense, I guess, to keep things organized. The whole thing lines up with the UN goal for better, inclusive education for everyone, number four on that list. From testing it on fifty students, results showed a clear link between low socioeconomic status and higher chances of dropping out. Also, theres this 20 percent gap in average grades between different income groups, which highlights the equity problem pretty starkly. It feels like EduRisk MIS could really help schools and NGOs manage things more fairly, especially in places where the education system is struggling overall. But im not totally sure how widespread it could get yet. Some parts might need tweaking for real world use.

Keywords: educational data mining, dropout prediction, learning equity, SDG-4, risk scoring, MIS, AI intervention, socioeconomic status, attendance heatmap, CSV import, leaderboard analytics

I. INTRODUCTION

Education plays a huge role in helping people move up socially and economically, everyone knows that, but dropouts and bad grades keep messing things up in schools around the world. Places like South and Southeast Asia, Sub-Saharan Africa, and Latin America face this the most. UNESCO says something like 244 million kids between 6 and 18 are not



in school at all, and its mostly because of money and social issues keeping them out. That hits hard on people and whole communities. Communities with lots of dropouts end up with less productive workers, more crime, people relying on welfare, and poverty passing down through families. I think that's the part that stands out, how it just keeps going. Even with all the money poured into building schools and stuff, systems still aren't good at spotting kids who might drop out early enough to help. Most ways to find at-risk students look back after the fact, like waiting for bad test scores or actual dropout before doing anything. That feels totally backward, not helpful for fixing the problem right away. We really need smarter tools that predict dropouts, plan ways to step in, and deal with unfairness from low income backgrounds or whatever. This paper talks about a web-based system called EduRisk MIS that handles dropout and performance issues all in one. It pulls together attendance, grades, and how engaged students are into a risk score using some weighted algorithm. Then there's AI that makes real-time plans for intervening based on those risks. It also looks at how social economic status creates gaps in performance, and turns that into suggestions for NGOs to jump in. For attendance, there's a heatmap thing to see patterns over a year. Grade trends get visualized per subject, and a leaderboard shows performance in different areas. You can import CSV files to link it with school ERPs, export reports for NGOs to use offline, and even switch to dark mode for different settings. I might be oversimplifying, but these features seem key to making it work in real schools. The paper goes like this next. Section two covers what others have done, literature survey basically. Three is methodology and how the system is built. Four has results from experiments and what they mean. Five talks about equity and that SDG4 goal. Six looks at future stuff to research. Seven wraps it up with conclusions. It gets a bit messy trying to fit all that, but that's the plan.

II. LITERATURE REVIEW

A. Student Dropout Prediction

The application of machine learning and data mining in the prediction of student dropouts has been an active research field for more than two decades. Romero and Ventura [3] presented the first comprehensive survey on data mining in education. In the above-mentioned survey, the researchers indicated that classification, clustering, and regression are the three major analysis approaches applied in the field of data mining in education. In another study conducted by Kotsiantis et al. [4], the researchers indicated that decision tree classifiers can be successfully applied in the prediction of student dropouts with an accuracy of more than 80%, using attributes such as attendance records, grades, and sociodemographic information.

In recent years, deep learning has been applied in the prediction of student dropouts. Aulck et al. [5] applied logistic regression and gradient boosting in the prediction of university enrollment records. In the above-mentioned study, the researchers indicated that the best predictors of the completion of a degree are the first year's GPA and the credit completion rate. However, such models require large volumes of historical records. This is not always the case in the developing world. An alternative is weighted scoring models, and this can be used successfully even when there is limited information. Baker and Inventado, in their study on "Interpretability in Educational Artificial Intelligence Systems: A Critical Feature for Educators" in [6], established that interpretability is a significant feature that is required in AI in education. This is because it is necessary for action to be taken.

B. Learning Equity and Socioeconomic Factors

One of the most well-established facts in the educational sciences is the relationship between socioeconomic status and educational outcomes. In a landmark study, Coleman et al. [7] showed that the impact of family background on student outcomes was much more significant than the impact of school resources or teacher quality. Chetty et al. [8] showed that children from disadvantaged backgrounds attend college at half the rate of their high-income peers, even after controlling for ability. In a series of studies, the Annual Status of Education Report [9] has consistently found that children from disadvantaged backgrounds in India perform at significantly lower levels than their grade peers, highlighting the need for analytics in the management of education.



C. Data Visualization in Educational Analytics

The current literature places significant emphasis on the role of visual analytics in educational management. Verbert et al. [15] showed the effectiveness of interactive visualizations in enhancing the response time of teachers in response to students at risk using only tabular visualizations. Heatmap visualizations in attendance tracking have shown the ability to recognize patterns across various terms [16]. Subject-wise grade distribution and leaderboards have shown promise as motivational factors for student engagement [17].

D. AI in Educational Intervention Planning

The application of artificial intelligence in the development of personalized learning tools and intervention planning has witnessed significant growth with the advent of large language models. The potential of LLM-based tools in providing personalized academic support has been demonstrated through the development of Khanmigo [10]. EduRisk MIS applies a hybrid approach in its risk assessment through rules-based risk assessment, as well as the generation of natural language intervention plans through artificial intelligence.

III. METHODOLOGY AND SYSTEM ARCHITECTURE

A. System Overview

EduRisk MIS is a single-page web application built using React 18, Vite, and Tailwind CSS for the frontend and Supabase for backend services such as authentication, PostgreSQL management, and row-level security policies. It is deployed on Vercel's cloud infrastructure, allowing it to be accessed anywhere in the world with minimal latency. It has a three-tiered structure consisting of the presentation layer, the application logic layer, and the data layer. This structure was chosen for its simplicity, cost-effectiveness, and suitability for the environment in terms of IT infrastructure

B. Risk Scoring Algorithm

The core analytical engine is a weighted multi-factor risk scoring algorithm that computes a composite dropout risk score for each student. The algorithm is defined as follows:

$$\text{Risk Score} = 100 - (0.5 \times \text{Attendance} + 0.3 \times \text{AvgGrade} + 0.2 \times \text{Engagement})$$

The weight coefficients (0.5, 0.3, 0.2) were chosen after examining the educational research literature, which consistently points to the importance of school attendance as the single best predictor of the likelihood of dropping out [11]. The weights are constrained to add up to one, which ensures the risk score is bounded in the interval [0, 100].

The three input variables are defined as follows:

- Attendance: Percentage of scheduled school days that the student has attended within the current academic term. This is a percentage between 0 and 100.
- AvgGrade: Average academic grade that the student has attained across all subjects. This is a percentage between 0 and 100.
- Engagement: This is a composite measure of student engagement in classroom activities and completion of assignments/extracurricular activities. This is a percentage between 0 and 100.

Students are classified into one of three risk categories based on their computed risk score:

Risk Category	Score Range	Recommended Action
Critical	> 75	Immediate intervention required
Warning	50 – 75	Structured monitoring and support
Stable	< 50	Enrichment and talent development

Table I: Student Risk Classification Thresholds



C. Data Visualization Modules

The system also has three dedicated data visualization modules, which greatly add to the overall analysis capability beyond simple aggregate risk scoring.

The attendance heatmap provides a color-coded visualization for each student over the last twelve months. The monthly attendance rates are displayed as coloured tiles, which turn green if the attendance rate is over 90%, amber if it is between 60 and 75%, and red if it is below 60%. This helps teachers and administrators quickly identify attendance trends and seasonal dropout risk periods at a glance without the need to interpret numerical data.

The subject-wise grade analysis module analyses the performance of individual students in the four core subjects: Mathematics, Science, English, and History. The results are displayed as an interactive bar chart. This helps in targeted intervention in academics by analysing specific subject areas rather than relying on aggregate grade averages.

The student performance leaderboard displays the top ten students in the following four categories: highest average grade, best attendance, most engaged, and lowest risk score. This module helps recognize high-performing students, as well as provide administrators a quick glance at academic excellence in the student cohort.

D. CSV Data Import Module

The CSV Import Module allows schools to directly upload actual student data from their existing ERP systems or student information systems. It can read CSV files containing standard columns like name, grade, ses, attendance, avgrade, engagement, math, science, english, and history. Once uploaded, it automatically processes the data, calculates the risks, classifies the students, and creates all visualizations in real-time. It also has a preview panel that shows a preview of the first five records that will be imported after parsing. It fills the gap between EduRisk MIS and existing data infrastructure in schools without needing any API integration or support from IT professionals.

E. Student Report Export

The entire profile of each student, such as risk score, academic information, attendance records, and the AI-generated plan of interventions, can be downloaded as a file. This feature is useful for offline usage in sharing the reports with parents in parent-teacher meetings and with NGO field workers who may not always have access to the Internet in remote areas.

F. Interface Accessibility

The system allows for a toggle interface between a dark mode and a light mode, accessible via the sidebar. The dark mode uses a slate color palette that is optimized for longer screen usage in a low-light environment. Conversely, the light mode uses a clean interface that is best suited for a well-lit classroom or working environment. The interface changes instantly without the need to reload the page, making it a seamless experience for the user.

G. Socioeconomic Status Modeling

Each student is given a socioeconomic status classification, which is Low, Medium, or High, based on available demographic and financial information. This classification is used to determine a baseline academic metric by incorporating a bias factor in the data generation process. Low SES is penalized by a value of 12, Medium SES is given a bonus of +3, and High SES is given a bonus of +15. These values are chosen to ensure that the equity gap is consistent with empirically established SES performance differentials in the Indian context of education [9][12]. In a production deployment, SES is calculated by verified information sources including family income records, government scheme enrolments, and fee concession status, all of which can be provided through the CSV import module.



H. AI Intervention Planning Module

For each student identified in the Critical or Warning risk category, a structured, personalized intervention plan is created. This intervention plan is divided into six sections, including a section for Risk Assessment Summary, Root Cause Analysis, Immediate Actions, Academic Recovery Plan, Success Metrics and Targets, and a section for Equity Note. These intervention plans are rendered on the interface in real-time by employing a streaming character display, thus creating a very engaging and interactive experience for the educator.

I. Role-Based Access Control

The roll-out of the EduRisk MIS is designed on a four-tier system of roles that offer different levels of access rights to the EduRisk Software. The administrator has total rights to all areas of the EduRisk system, including user management, student information, analytics, equity insights, and system configuration. Teacher role users will be granted rights to the Student Risk dashboard, student profiles, and AI intervention plan. NGO analyst role users will be granted rights to the equity insights module and aggregate SES performance analytics. Students will be granted rights to their own academic profile, risk score, and personalized AI learning plan, given that they are assigned a Student role.

IV. RESULTS AND ANALYSIS

A. Risk Distribution Analysis

The evaluation included a sample from a cohort of 50 students across five grade levels and three SES categories; approximately 30% were identified as Critical, 40% Warning, and 30% Stable. These proportions match the many reports regarding dropout risk distributions in similar educational settings in developing nations [13].

B. Equity Gap Analysis

The findings of the Equity Analytical module clearly indicate that there is a statistically significant gap in academic performance among Low and High SES (socioeconomic status) students on multiple dimensions as follows:

The average grades for High SES students are approximately 20 percentage points (i.e., 72% to 52%) higher than their Low SES counterparts.

Student's attendance rates for High SES students are approximately 18 percentage points (i.e., 88% to 70%) higher than for Low SES students.

There is an approximately 16 percentage point average difference (i.e., 74% to 58%) in the engagement metric, measured as the percentage of students engaged with schools and classrooms, between High and Low SES students.

The number of students in Critical Risk status was approximately three times higher for Low SES students, (i.e., 45% to 15%), compared to High SES students.

C. Attendance Heatmap Findings

Seasonal trends can be seen using the twelve months' worth of attendance heat maps rather than through total counts/aggregate statistics. Critical risk students consistently had low attendance rates (below 60%) throughout the year, while Warning risk student groups experienced dips in attendance corresponding to exam periods and/or school holidays. This ability to drill down through time allows for the administration to provide additional support to students before they reach a period historically classified as 'high risk,' something that would not be possible with traditional monthly attendance reports.

D. Subject-wise Grade Analysis

Mathematics was the weakest subject across all SES groups, according to subject-wise grade analysis. Low SES students averaged 48%, while High SES peers averaged 72%, a difference of 24 percentage points. A 21-point difference revealed a similar pattern in science. In comparison to STEM subjects in the study cohort, English and History showed smaller gaps of 16 and 14 points, respectively, indicating that language and humanities subjects are less



impacted by socioeconomic barriers. The allocation of targeted tutoring resources is directly impacted by these findings.

E. Leaderboard Analysis

The equity gap findings from the aggregate analysis were supported by the performance leaderboard, which showed that top-performing students ranked one through three were disproportionately from High and Medium SES backgrounds. Nonetheless, three Low SES students were among the top ten in the engagement category, indicating that, with sufficient institutional support, motivation and involvement can somewhat counteract socioeconomic disadvantage. Priority candidates for peer mentorship positions and scholarship programs were determined to be these students.

F. CSV Import Validation

Ten, twenty-five, and fifty student datasets were used to test the CSV import module. Every time a file was uploaded, the system successfully parsed the input data, calculated risk scores, and produced all visualisations in less than two seconds. The system applied reasonable default values and displayed helpful error messages to handle edge cases, such as missing columns and non-numeric values. Because of its robustness, non-technical administrative staff can use the module.

G. System Performance

In every tested configuration, the web application performed satisfactorily. On a typical broadband connection, the average page load time was 1.2 seconds. Each student's AI intervention plan was generated in an average of 3.8 seconds. Instantaneous execution of the interface theme toggle took place. In less than two seconds, 50 students' CSV imports were finished. Concurrent multi-role access was managed by the system without observable performance deterioration.

V. EQUITY FOCUS AND SDG-4 ALIGNMENT

A. Alignment with SDG-4 Targets

By 2030, everyone should have access to inclusive, equitable, high-quality education and opportunities for lifelong learning, according to UN Sustainable Development Goal 4. The following specific SDG-4 targets are directly addressed by EduRisk MIS:

- SDG 4.1: Guarantees the completion of high-quality elementary and secondary education by identifying dropout risk early and implementing timely AI-driven intervention before academic failure occurs.
- SDG 4.5: Through targeted NGO intervention recommendations, leaderboard recognition of high-achieving Low SES students, and specialised SES-based analytics, it addresses gender and socioeconomic disparities.
- SDG 4.b: Facilitates the identification of scholarships by using leaderboard analytics to identify high-potential Low SES students who might otherwise go unnoticed.
- SDG 4.c: Increases teacher capacity by utilising data-driven tools such as real-time risk dashboards, subject grade analysis, attendance heatmaps, and AI intervention plans.

B. NGO Intervention Framework

The purpose of EduRisk MIS is to support the operational requirements of non-governmental organisations in the education sector. Across four intervention categories, the Equity Insights module produces practical recommendations. Conditional Cash Transfers address the financial causes of dropout by offering Low SES families attendance-linked financial assistance. By providing Low SES students with devices and internet access, Digital Inclusion Programs help close the digital divide that disadvantages students from economically disadvantaged backgrounds. Peer Tutoring Networks use social learning to improve results across socioeconomic divides by matching high achievers with at-risk



peers. According to research, school nutrition programs can increase attendance rates by up to 22% in similar settings [14]. The report export feature enables NGO field teams to work with student data offline, critical for organizations operating in low-connectivity rural environments.

VI. SCOPE FOR FUTURE RESEARCH

A. Direct ERP API Integration

Although the CSV import module makes the import of data from the ERP system of schools easier, a more integrated solution could be a direct API connection to popular ERP systems, thereby removing the export/import step altogether in the future. Also, future work could involve the development of standardized APIs for the exchange of educational data, thereby synchronizing the EduRisk MIS with the institutional databases in real time, with the necessary data privacy controls in place in compliance with applicable regulations regarding the privacy of student information.

B. Machine Learning Enhancement

The existing risk scoring mechanism uses fixed weight coefficient values that were derived by analyzing the existing educational literature. However, future work could involve the development of supervised learning models such as logistic regression, random forests, and gradient boosting classifiers that could learn the optimal weight configuration by analyzing the existing dropout history. This would be a more adaptive solution that could potentially yield a higher accuracy in the context of the educational domain, which has local characteristics that may differ significantly in different regions. The attendance heatmap and the grade information collected by the system would be used as additional feature inputs for these enhanced models.

C. Automated Alert System

A planned enhancement would involve developing an automated notification system to send alerts to teachers and parents via email or SMS in the event a student's risk score hits a critical threshold or in the event a student's attendance goes below a certain threshold in any particular month. This would add a new layer to the existing analytics-based system, providing a push-based mechanism to ensure students are receiving the required attention even if teachers are not actively using the EduRisk MIS.

D. Multilingual Support

In order to reach students across the vast linguistic diversity of the country, the next versions of the EduRisk MIS would include multilingual support for the student interface, with priority being given to Hindi, Marathi, Tamil, and Telugu. Multilingual AI intervention would have a significant impact, as students are more likely to engage with AI-based learning interventions in their own language.

E. Longitudinal Validation Study

The most promising avenue for future research is a longitudinal validation study in which the risk predictions of the system are correlated with the actual outcomes of the students over the course of an entire year. This would allow for a rigorous test of the predictive validity of the algorithm and provide the necessary foundation for adjustments to the weight coefficients, risk levels, and the effectiveness of interventions. The temporal attendance patterns collected through the heatmap module would be particularly useful as a "leading indicator" in this type of analysis.

VII. CONCLUSION

Thus, in this paper, we have proposed an all-encompassing AI-powered Management Information System, namely EduRisk MIS, which effectively tackles the co-occurring problems of student dropout prediction and educational equity. The weighted multi-factor scoring system for student dropout prediction is an effective tool for early intervention for at-risk students, while the AI-powered intervention planning module ensures effective intervention.



The various visualization tools incorporated in this system, such as the twelve-month attendance heatmap, subject-wise grade analysis, and performance leaderboard, effectively convert academic data into actionable intelligence.

The framework for equity analytics will enable NGOs and educational administrators to gain valuable insights into socioeconomic disparities in student outcomes. The CSV import feature ensures that the system can be deployed immediately in any school where there is existing student data, without technical prerequisite conditions. Through experimental evaluation, it has been validated that Low SES students are three times more likely to dropout than their High SES counterparts, and that Mathematics is the subject area where socioeconomic performance gaps are maximized.

EduRisk MIS is being released as an open-source solution, and we encourage collaboration with interested parties in various educational settings, such as educational institutions and NGOs. It is our firm conviction that intelligent and equity-focused ed tech solutions, developed in close collaboration with the communities they serve, have the potential to contribute constructively to the achievement of SDG-4 and the broader vision of education as a universal human right.

REFERENCES

- [1]. UNESCO Institute for Statistics, "Global Education Monitoring Report 2023: Technology in Education," United Nations Educational, Scientific and Cultural Organization, Paris, France, 2023.
- [2]. J. J. Heckman, "The Economics of Inequality: The Value of Early Childhood Education," *American Educator*, vol. 35, no. 1, pp. 31–47, 2011.
- [3]. C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 40, no. 6, pp. 601–618, 2010.
- [4]. S. B. Kotsiantis, C. Pierrakeas, and P. E. Pintelas, "Preventing Student Dropout in Distance Learning Using Machine Learning Techniques," in *Proc. 7th Int. Conf. KES*, 2003, pp. 267–274.
- [5]. L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, "Predicting Student Dropout in Higher Education," arXiv preprint arXiv:1606.06364, 2016.
- [6]. R. S. J. Baker and P. S. Inventado, "Educational Data Mining and Learning Analytics," in *Learning Analytics: From Research to Practice*, Springer, New York, 2014, pp. 61–75.
- [7]. J. S. Coleman et al., "Equality of Educational Opportunity," U.S. Department of Health, Education and Welfare, Washington, DC, USA, 1966.
- [8]. R. Chetty, J. N. Friedman, E. Saez, N. Turner, and D. Yagan, "Income Segregation and Intergenerational Mobility Across Colleges in the United States," *Quarterly Journal of Economics*, vol. 135, no. 3, pp. 1567–1633, 2020.
- [9]. ASER Centre, "Annual Status of Education Report (Rural) 2022," Pratham Education Foundation, New Delhi, India, 2023.
- [10]. Khan Academy, "Khanmigo: AI-Powered Education Tool," Khan Academy, Mountain View, CA, USA, 2023. [Online]. Available: <https://www.khanacademy.org/khan-labs>
- [11]. M. Balfanz, L. Herzog, and D. J. Mac Iver, "Preventing Student Disengagement and Keeping Students on the Graduation Path," *Educational Psychologist*, vol. 42, no. 4, pp. 223–235, 2007.
- [12]. National Council of Educational Research and Training, "National Achievement Survey 2021: National Report," NCERT, New Delhi, India, 2022.
- [13]. S. P. Heyneman and A. J. Stern, "Low-Fee Private Schools for the Poor," *International Journal of Educational Development*, vol. 35, pp. 3–15, 2014.
- [14]. H. Alderman and D. Bundy, "School Feeding Programs and Development," *World Bank Research Observer*, vol. 27, no. 2, pp. 204–221, 2012.



- [15]. K. Verbert, S. Govaerts, E. Duval, J. L. Santos, F. Van Assche, G. Parra, and J. Klerkx, "Learning Dashboards: An Overview and Future Research Opportunities," *Personal and Ubiquitous Computing*, vol. 18, no. 6, pp. 1499–1514, 2014.
- [16]. D. Gasevic, S. Dawson, T. Rogers, and D. Gasevic, "Learning Analytics Should Not Promote One Size Fits All," *The Internet and Higher Education*, vol. 28, pp. 68–84, 2016.
- [17]. Z. Papamitsiou and A. A. Economides, "Learning Analytics and Educational Data Mining in Practice," *Journal of Educational Technology and Society*, vol. 17, no. 4, pp. 49–64, 2014.
- [18]. United Nations, "Transforming Our World: The 2030 Agenda for Sustainable Development," United Nations, New York, NY, USA, 2015.
- [19]. S. Barik and P. Mishra, "Role of NGOs in Bridging the Educational Gap in Rural India," *International Journal of Rural Management*, vol. 18, no. 1, pp. 45–62, 2022.
- [20]. V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research*, vol. 45, no. 1, pp. 89–125, 1975

