# Educational Data Mining for Dropout Risk Prediction

**Tresna Dey**

## Abstract

*The seat in the colorful classroom of the Sylhet secondary school, or the overcrowded school of Dhaka, is a narrative of potential squandered in a madrasah of Barisal rural or the crowded schools of Dhaka. Although Bangladesh has recorded notable advances in its primary education system wherein it has nearly one hundred percent enrollment in primary schools, the secondary education system has been a mountainous and bumpy ride.*

According to government surveys as of the close of 2025, there is an ugly truth about the situation in the country: only 84 percent of children have finished their primary school, with merely 43.9 percent managing to complete their secondary education. Such a sharp contrast underlines that access is not good enough but retention, quality, and support systems are as well important. Every student who drops out is a loss to the individual student and to the society, a missed opportunity, economically-productive potential, social mobility, and human developmental pathway of the country.

It is no longer, why do students drop out, but who is likely to drop out next and when. With the dawn of the so-called Educational Data Mining (EDM) and the latest wave of powerful machine learning (ML) algorithms, researchers and policymakers can now transition away, on the one hand, to retrospective statistics and, on the other hand, to proactive intervention. Using both historical and real-time data, predictive models are able to point out at-risk students prior to their total deengagement. It is a paradigm change: dropout is no longer an effective metric but a predictable phenomenon and schools, communities and governments are capable of intervening in a timely, planned and fair manner.

The research on educational data mining has shown that it is not a single situation that leads to dropout; rather, it is a complex convergence of socio-economic, academic and behavioral risks. Random Forest, XGBoost, and Gradient Boosting machine learning models have been used on datasets with more than 8000 students spread throughout Bangladesh and have identified high impact predictors frequently missed by traditional statistical techniques. The socio-economic indicators are always among the most powerful

indicators. Households that have less than a monthly income of 117.71 (around 14000 BDT) have over ten times the probability of dropping out than those earning more than that. But poverty is no more than half the story. Education especially the level of education attained by the mother is a good protective factor. The odds ratio of students whose mothers are educationally attained shows a significant reduction on the likelihood of dropping out, indicating the inter-generational positive effects of maternal literacy of 2.8 odds ratios.

The risk of dropout is also determined by geography and access. Residing over half a kilometer to school like in rural and peri-urban areas adds to the chances of non-continuation, especially through girls. Distance in itself might appear insignificant, but the distance, in combination with various contextual variables: safety issues, transport expenses, and socio-cultural beliefs regarding female movement. These dangers are further increased by the so-called rickshaw cost factor and the unwillingness of parents to have their girls travel alone. These results demonstrate that the socio-economic and spatial vulnerability are not independent of each other and these models must have the ability to predict using multiple dimensions, rather than in isolation.

Another early warning system is academic performance. One of the most immediate indicators is attendance patterns: students that have a declining attendance below 80% in one term have a much greater likelihood of permanently disengaging. Likewise, problem-specific conflicts, especially in Mathematics and Science, are highly associated with the risk of dropout in the long run. As a case in point, students who showed weaknesses in these core subjects by Class 8 are 40% likely to drop out of school by Class 10. These trends highlight the fact that early academic interventions can disproportionately affect, as well as that predictive analytics needs to be based on granular performance, and not just aggregate grades or end-of-test results.

This data is converted into actionable information by the use of predictive modeling. Using classification algorithms, schools may assign students a customized set of risk scores, a measure of the likelihood of the student dropping out due to a multidimensional system of 24 socio-economic, behavioral, and academic factors. Random Forest Regressors have demonstrated a high level of accuracy in the Bangladesh case with an accuracy of 0.973 (R2) which can be used to identify high-risk students with high confidence. Most importantly, these models currently have fairness measures to counter bias, and predictions are now made in a fair manner, without regard to gender, geography, and socio-economic status. Practically, it allows the algorithms to avoid unintentionally punishing students due to systemic disadvantages which may be poverty or living in rural areas and instead allows the interventions to reach the areas most in need.

Social consequences of dropping out go well out of the classroom. Loss of each year of education is associated with lower household income, less labor productive work force and low Human Development Index (HDI) scores. As a result, dropout prevention is not only a significant educational issue, but also an

economic and social necessity. Interventions that are based on EDM are therefore consistent with national development priorities as they provide a data-driven solution to poverty and inequality cycles.

The real transformative capability of such AI systems is revealed once the systems are combined with social safety nets and support mechanisms at the school level. To illustrate, predictor models make it possible to allocate financial support. Instead of issuing standard stipends, the government can allocate resources to the students that have been identified as having the highest score of Financing Risk, as this way the impact of limited funds will be the most. On the same note, predictive alerts may support parental involvement: automated SMS messages or visiting the community by volunteers can solve barriers before they arise, like transportation barriers, family obligations, or health-related barriers, and minor cases may turn into long-term disengagement.

AI can also be used to the advantage of teachers. Predictive dashboards are used as digital assistants in classrooms with student-teacher ratios that are likely to be overwhelming. Teachers are provided with practical information about those students who need additional support in the gatekeeper courses such as Math and English as they are able to implement specific interventions and avoid the general ones. The strategy lowers the workload of teaching professionals, enhances the effectiveness of retention and assists in making interventions timely and individual.

The ethical and policy aspect of the data use must also be considered when the implementation of EDM in Bangladesh is implemented. The privacy of students should be ensured, and the personal and performance data gathering should be conducted according to the sound frameworks of consent. Models should be transparent and interpretable in order to avoid the black box effect, where the educators and policymakers cannot interpret or trust the predictions. Along with this, ethical focus on algorithmic fairness, transparency, and accountability is critical in 2026, which will make AI a provider of equity and not a tool of strengthening system inequalities.

Besides, predictive analytics integration in a wider-policy effort increases influence. EMD observations can be used to guide national policies regarding school building, staffing, and the curriculum. As an example, when groups of high risk students are in specific geographic areas, policy makers have the opportunity to focus on resources allocation, such as transport subsidies, school expansions or special teacher training programs. Equally, when the curriculum designers are aware of the subject-specific risk pattern, they can introduce remedial courses, after-school tutoring, and peer-mentoring initiatives in a data-driven and economical fashion.

This is a vital predictive ecosystem that requires community involvement. Involvement of parents, local leaders, and local organizations is a way of making sure that interventions are culturally relevant and contextual. Predictive alerting combined with community-based follow-ups can make the difference

between an action being taken and the intended result, and as such, a cycle of data-driven action and model-driven data can be established. This makes this cycle more predictive as well as the social validity of AI interventions.

To conclude, student dropout is not unavoidable and arbitrary. It is a complicated, foreseeable phenomenon that is formed as a result of the interaction of social-economic, academic, and behavioral components. Educational Data Mining includes the means to foresee the risk, to focus the intervention, and to help students before they lose interest permanently. Such predictive systems in the vision of Smart Bangladesh, are not a tool of surveillance, but an umbrella to the vulnerable, so that technology is an enabler of opportunity, and not a controller of inequality. By detecting the hidden indicators of distress, e.g., attendance declines, lack of performance in particular cases, economic instability, and physical remoteness, schools and governments and communities can intervene without hesitation to ensure students stay involved. Every occupied desk is an individual lives pathway saved, a family bolstered and a society built up.

Through the incorporation of predictive analytics and social safety nets, teacher support system, parental engagement, and ethical governance, Bangladesh would be able to change the current state of secondary education management in the country towards proactive stewardship. By so doing, the country will transform the vicious cycle of dropout into the virtuous cycle of education, opportunity, and social mobility. Educational Data Mining has ceased being a device of wonderment and is now becoming a tactical device to determine the future of millions of students with all the children having a chance of finishing their secondary schooling and contributing to the social and economic life of the nation in all aspects. By so doing, AI and predictive modeling is not simply a technological advancement, but an equity, inclusion, and national development stimulus.