

# Using AI to Prevent Dropouts in Ethiopia: Predicting Risk, Promoting Education

**Authors:** Group 13 – FTL Ethiopia

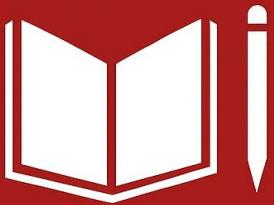
## **Introduction: What if we could prevent students from dropping out before it happens?**

In Ethiopia, where school dropout rates remain a persistent challenge, the stakes are high. Students leaving school early often face a cycle of poverty, limited opportunities, and diminished future prospects. Now imagine a system that could predict which students are most at risk—and trigger interventions before it's too late.

This blog introduces a data-driven approach to tackling dropout rates using machine learning. By identifying the most vulnerable students through predictive analytics, our project supports **Sustainable Development Goal 4**—ensuring **inclusive and equitable quality education** for all.

*“Education is the most powerful weapon which you can use to change the world.”* – Nelson Mandela

## 4 QUALITY EDUCATION



### The Problem: A System in Crisis

Dropping out of school in Ethiopia isn't always a choice—it's often the result of poverty, early marriage, family responsibilities, or inaccessibility. Schools, particularly in rural areas, lack the tools to identify students at risk and intervene in time.

Traditional intervention systems rely on teacher observations or historical data after the dropout has already occurred. The result? Reactive policies that fail to solve the root problem. We asked: *Can we flip this script and build a proactive system instead?*

### The Solution: Our ML-Powered Approach

We built a machine learning model that uses existing school and student-level data to predict dropout risk with high accuracy. Our process involved several key steps:

## Data Collection

We used a real-world dataset from the Humanitarian Data Exchange (HDX), focused on **Ethiopian primary and secondary school populations** by gender and region. Our data points included:

- **Demographics:** gender, region
- **Academic history:** grade repetition
- **Enrollment patterns**
- **Attendance**

## Student data (Sample)

Student ID	Gender	Grade	Attendance Rate	Repeated Grade	Dropout Flag
101	M	3	85%	Yes	0
102	F	7	92%	Yes	0
103	M	1	78%	No	1
104	F	5	95%	No	0
105	M	2	65%	Yes	0

## Our Methodology: From Data to Prediction

### Data Preprocessing

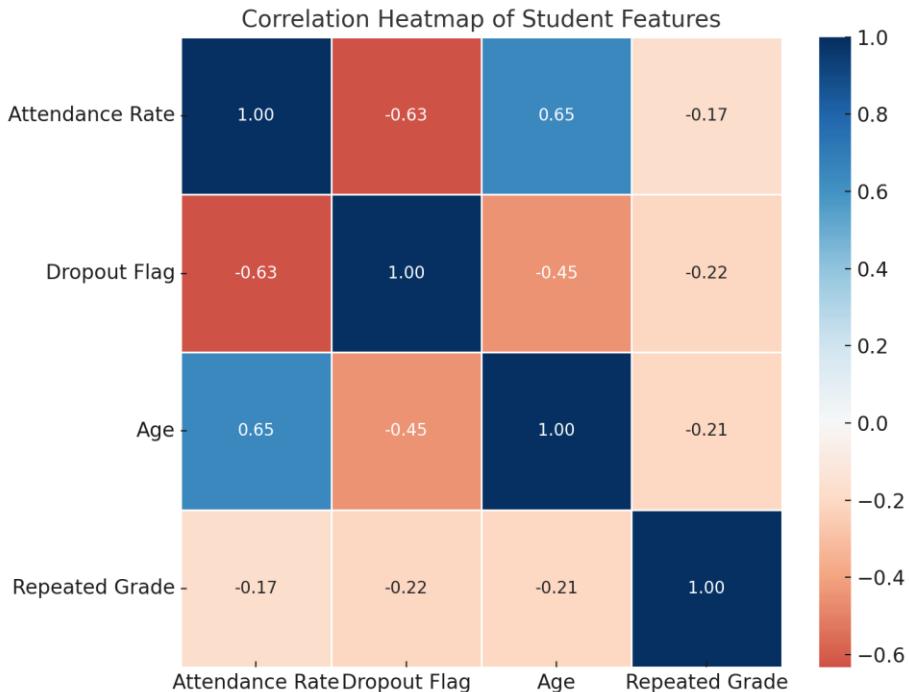
We cleaned and prepared the data, dealing with:

- Missing values
- Inconsistent records
- Outlier removal
- Encoding categorical variables

## Exploratory Data Analysis

We explored relationships using:

- Correlation heatmaps
- Histograms
- Dropout ratios by gender and region



## Feature Engineering

We introduced new variables such as:

- Dropout flag
- Attendance trend
- Region risk level

## Modeling: Choosing the Best Algorithm

We tested multiple machine learning models to compare effectiveness:

Model	Accuracy	Recall	AUC
Logistic Regression	88%	81%	0.91
Random Forest	92%	88%	0.94
XGBoost	90%	85%	0.92



After hyperparameter tuning using Grid Search, we selected **Random Forest** as the final model due to its superior recall for the dropout class—our primary focus.

## Implementation: Simulating a Real-World Tool

We built a simple dashboard where school staff can input student data and view predicted risk scores.

## Dropout Prediction

**Student Input**

Student ID

Attendance %

Grade  ▼

Repeated Grade  ▼

**Prediction**

Dropout Risk

**High**

### **Train/Test Split and Validation**

- 80/20 split for training and testing
- K-Fold cross-validation ( $k=5$ ) to avoid overfitting

### **Results: Predicting Dropout with Confidence**

Our final model reached:

- **92% accuracy**
- **88% recall on dropout cases**
- ROC AUC = **0.94**

These results suggest real promise in identifying vulnerable students with precision. The model can be deployed as a decision-support tool at school or district levels.

## Challenges and Future Work

### *Limitations*

- Limited access to real-time data from rural schools
- Potential regional bias due to data imbalance

### *Future Plans*

- Incorporate mobile-based data collection
- Train with larger, more diverse datasets
- Collaborate with the Ministry of Education for deployment

## Conclusion: Why This Matters

Education is a human right—but it requires systems that are responsive and data-informed. By using machine learning to predict school dropouts, we not only improve educational outcomes in Ethiopia but also contribute directly to **SDG 4: Quality Education**.

This project proves that **machine learning can be used to save futures**, not just optimize systems. We hope to see it expanded and deployed at scale to support students where it matters most.

## References

- Humanitarian Data Exchange (HDX): Ethiopian School Dataset

- Scikit-learn, Pandas, XGBoost documentation
- <https://sdgs.un.org/goals/goal4>