School Dropout Risk Analysis

Using Machine Learning to Predict and Prevent Student Dropout

1. Problem Statement

Early school dropout affects long-term academic and career success.
This project uses data analytics to identify students at risk and recommends actionable interventions based on real patterns.

2. Dataset Overview

Field	Details
Source	UCI Student Performance Dataset
Size	395 students (math stream)
Attributes	Grades (G1, G2, G3), study time, absences, parental and personal background

3. Target Definition

- A student is labeled **"At Risk"** if G3 < 10
- **33%** of students fall into this category

4. Exploratory Insights

- 📉 **Absences**: High-risk students have more absences
- 📘 **Study Time**: More study time = lower risk
- 👩 **Mother's Education**: Higher education = lower dropout probability

5. Predictive Modeling

Model Used: Random Forest Classifier
Accuracy Achieved: ~80%
Top Influencing Features:

- Past class failures
- Number of absences
- Study time
- Mother's education
- Internet access

6. Recommendations

1. **Track Absenteeism**
   Real-time alerts for abnormal attendance
2. **Provide Academic Support**
   Focused help for students with low study time
3. **Engage Parents**
   Especially where maternal education is low
4. **Early Warning System**
   Deploy ML tools in schools to proactively identify at-risk students

7. Conclusion

Machine learning enables schools to take proactive, data-driven actions.
With the right interventions, dropout rates can be reduced, improving student retention and academic success.