Identifying At-Risk Students with Classification

Presented to School Administrators

By: Jerome

Project Overview

- Goal: Build a classification model to identify students at risk of failing.
- Data includes academic performance, behavior, and demographic attributes.
- Classification allows early identification of students needing support.

📌Business Problem & Data Understanding

- Problem: Identify students likely to fail the final grade (G3 < 10).
- Data: Student records including study time, absences, grades, family background, etc.
- Target: Binary classification – Pass vs. Fail.
- Stakeholders: School management and academic support teams.

⚙ Modeling Approach

- Used classification models: Logistic Regression, SVM, and Random Forest.
- Input features: midterm grades, study habits, attendance, etc.
- Outcome: Predict whether a student is likely to fail the final exam.

Model Evaluation

- Measured performance using accuracy, precision, recall, and F1 score.
- Random Forest showed the best balance of performance and interpretability.
- Top predictors: previous grades (G1, G2), study time, and absences.

✓ Recommendations

- Use Random Forest model to flag at-risk students early in the semester.
- Provide targeted academic support based on predictions.
- Focus on students with low midterm scores and high absences.

🚀Next Steps

- Integrate model into student monitoring systems.
- Train teachers and staff to interpret model outputs.
- Expand dataset with more recent and behavioral data to improve accuracy.

Questions? Let's work together to support every student.

THANK YOU