Early Prediction of Student Dropout Risk

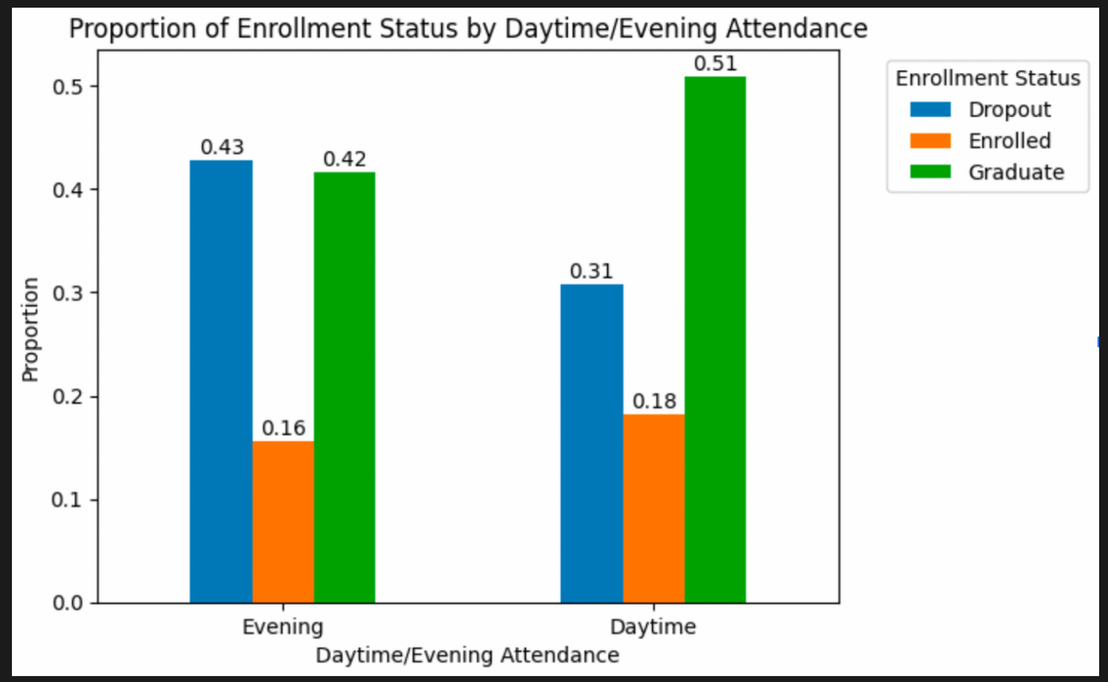
Objective

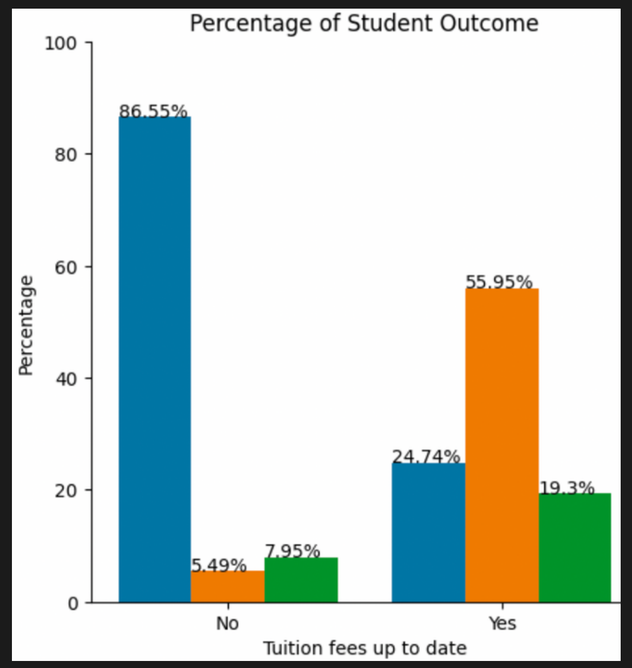
- Predict early risk of academic dropout among college students
- Identify at-risk students to implement supportive strategies

Dataset Description

- Instance: student
- Variables (37):
  - Academic paths
  - Demographics
  - Socio-economic factors
- Target: 3 categorical classifications (dropout, enrolled, and graduated)
- Data Source: retrieved from educational institutions in Portugal.

EDA Highlights





- Students attending daytime courses are less likely to dropout.
- Students unable to pay tuitions up to date are more likely to drop out.

Presented by Michelle Lin

Early Prediction of Student Dropout Risk

Feature Engineering

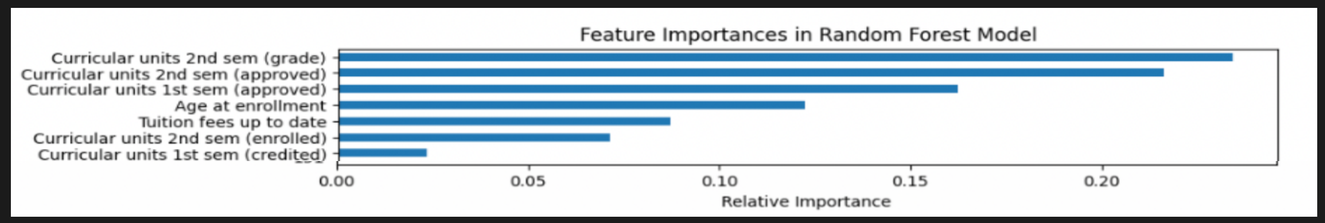
- Target Consolidation: Combined 'Enrolled' and 'Graduate' into one category.
- One-Hot Encoding: Applied to all categorical variables.
- Standardization: All numerical variables standardized to range 0-1.
- Feature Selection: Utilized Logistic Regression to identify the top 10 impactful features.

Key Predictors of Dropout

- Approved curricular units: Optimize curricular arrangement to ensure students can enroll in desired courses.
- Age at enrollment: Implement tailored support systems for older students.
- Tuition fees status: Offer scholarships to studnets in financial need.

Statistical Models

Model	Accuracy
Logistic Regression	0.7525
Decision Tree	0.8350
Random Forest	0.8520



Future Directions

- Expand research to inlcude international data for broader insights.
- Collect addtional variables like employment history.

Presented by Michelle Lin