

Using LMS Activity to Identify Early Academic Difficulty in First-Year Students

Chris Steadman, Caitlin Mills, Stephen Hutt

Background

Student dropout remains a problem in higher education, especially for first year students

Failure is often considered in binary terms which can obscure behavioral patterns for students most at risk

Defining failure: Final course grade below C-

Failure Groups:

- Multiple Failure (N=160)
- Single Failure (N=321)
- Non-Failure (N=4613)

Research Questions

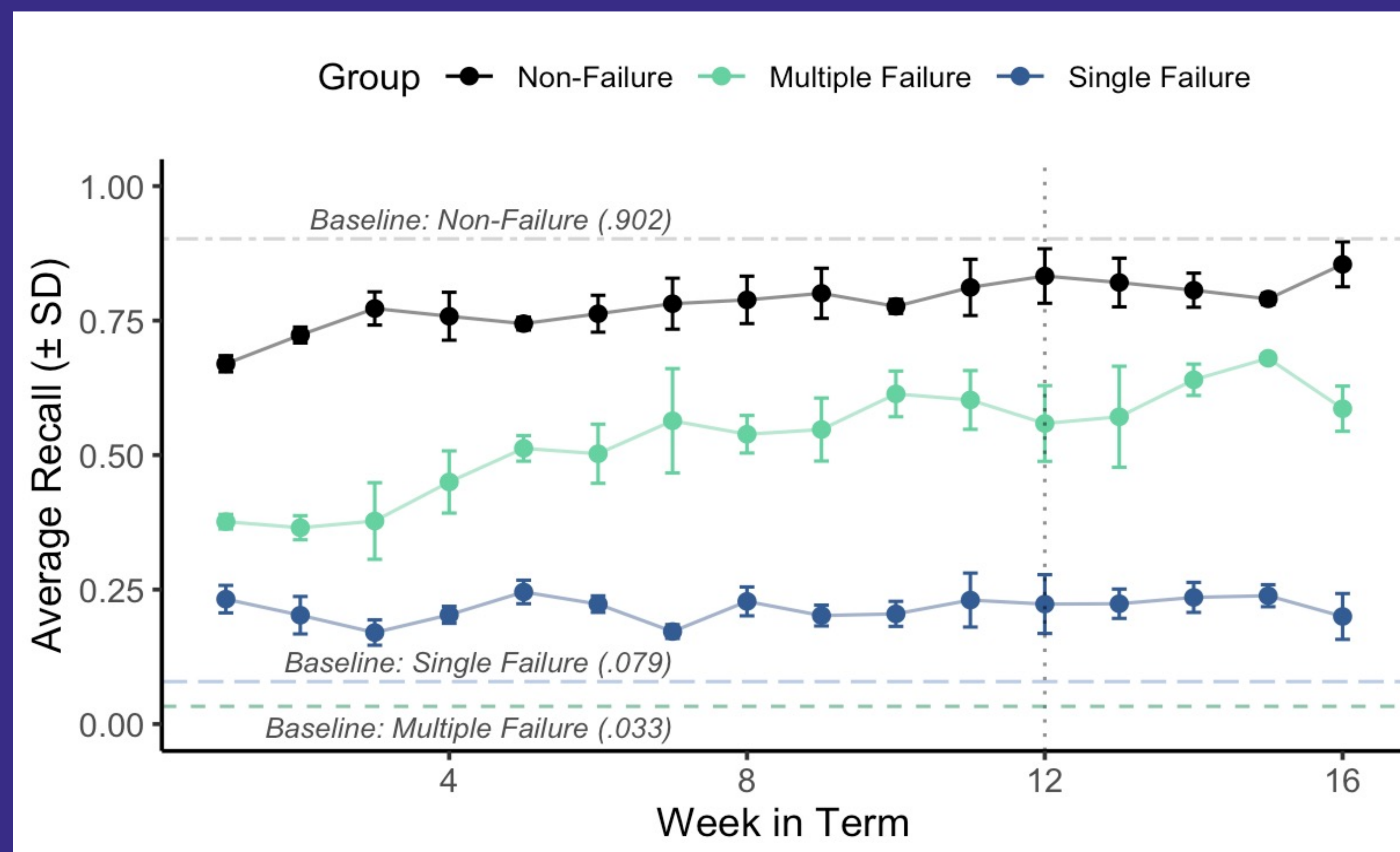
RQ1: Can we *accurately predict* that students will experience multiple failures in their first semester of college?

RQ2: How *early* can we make meaningful predictions about students' failure rate?

Student Demographics

	Overall (N = 5094)
Sex	
Female	54.7%
Male	45.2%
No Information Reported	0.2%
Race/Ethnicity	
Alaskan/American Native	1.5%
Asian	20.2%
Black/African American	9.2%
Hawaiian/Pacific Islander	0.2%
White	65.3%
No Information Reported	3.6%

Our model accurately predicted **over 50%** of students facing multiple failures at one month—and **68%** by finals.



Data Cleaning

Initial Data from Unizin Query (N = 8382)

Removed Learners:

- Missing Grade Data (N=429)
- Enrolled in less than 3 courses (N=1896)
- With Interactions before Fall 2022 (N=613)

Final Sample Data (N=5094)

Machine Learning Pipeline

Training/Testing Split: 80/20

Cross-Validation: 5-fold with randomized search across five unique data splits

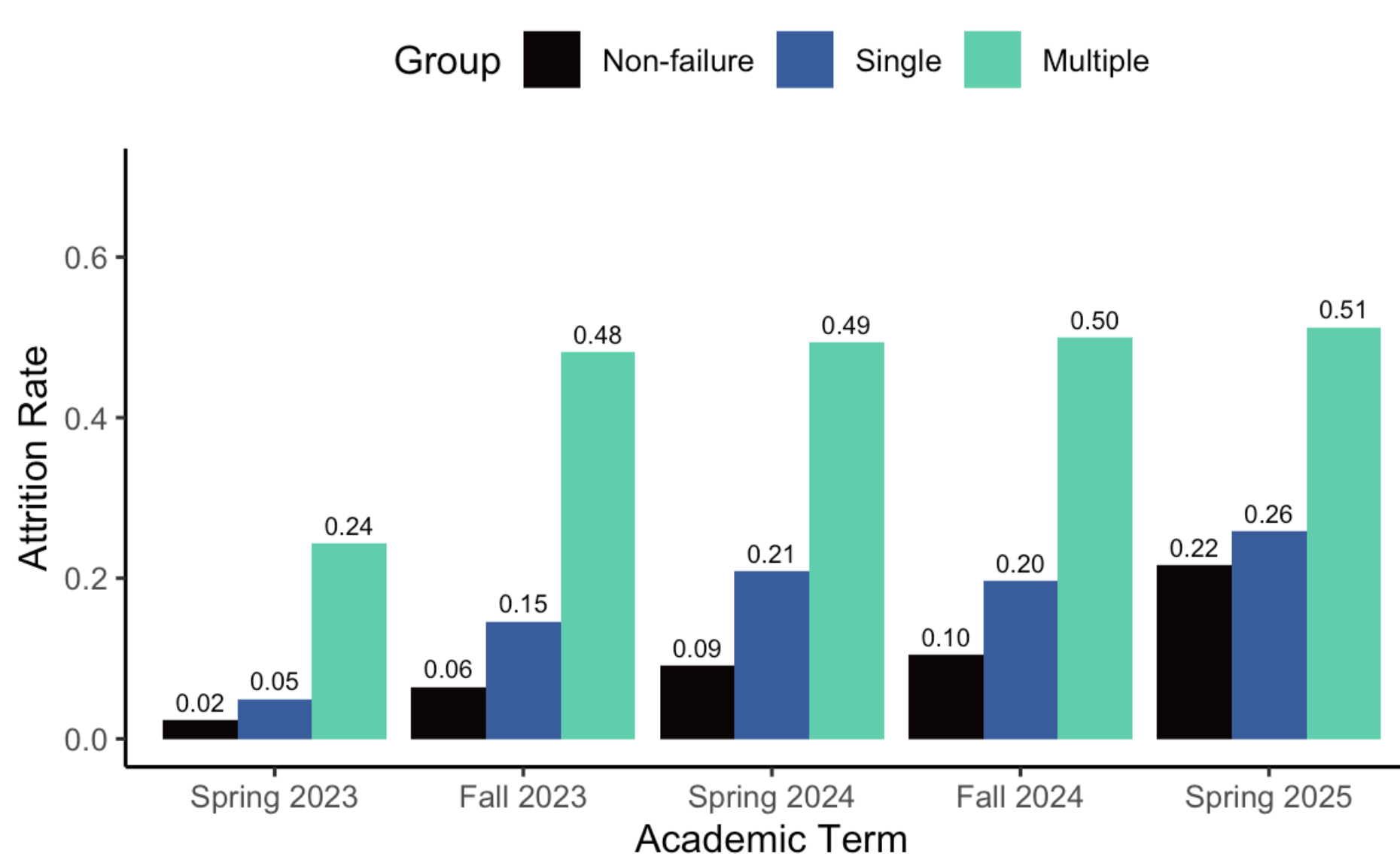
Model selection: Random Forest and XGBoost

Feature Engineering

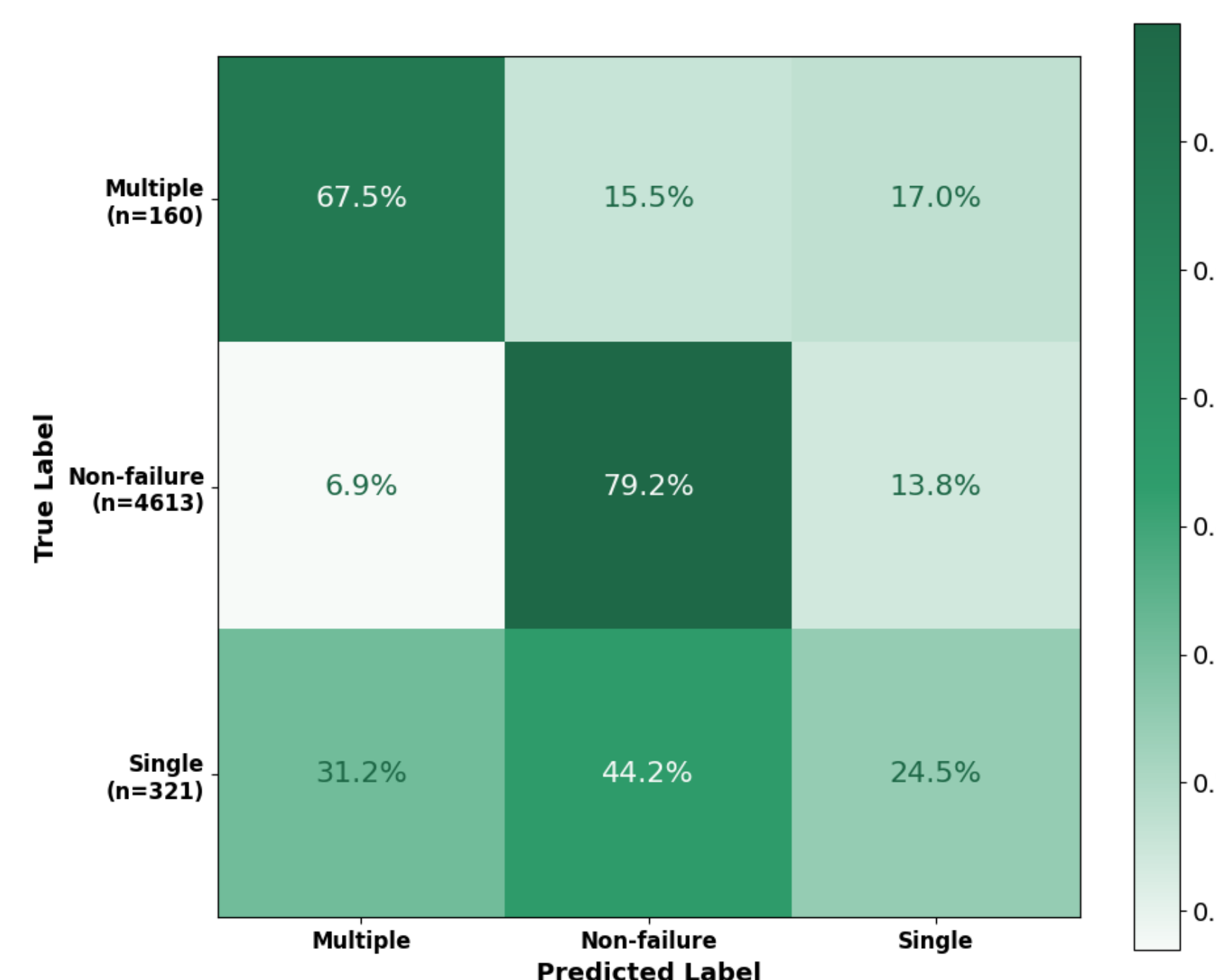
Feature Name	Description
Time-Binned Interaction Metrics	
Interaction Counts (By Time Block)	The raw count of interactions occurring during specific windows: Night (12 AM–6 AM), Morning (6 AM–12 PM), Afternoon (12 PM–6 PM), and Evening (6 PM–12 AM).
Interaction Proportions (By Time Block)	The proportion of the week's total interactions that occurred during each of the specific time windows listed above.
Volume & Intensity Metrics	
Average Weekly Time	Total active duration (in seconds) within Canvas
Average Weekly Actions	Frequency of specific purposeful actions
Average Number of Weekly 30-Minute Sessions	The count of distinct 30-minute blocks during which the student was active
Self-Regulation & Performance Metrics	
Total To-Do List Engagement	The raw count of interactions with the Canvas "To-Do" list or calendar.
Proportion of To-Do List Engagement	The ratio of To-Do list interactions to the total number of To-Do list interactions in the semester.
Average Weekly Submission Time	The average time elapsed (in hours) between the student's submission timestamp and the assignment deadline.
Average Student Activity Score	A composite weekly activity metric created by Unizin

Number of Features Each Week (k):
Total Features = (15×k) + (7×k)

Attrition Rate by Failure Group



Week 15 Confusion Matrix



Week 15 Feature Importance

