CMPT 310 — Milestone 1 Progress Report

Group: ClassPass
Members: Phil Akagu-Jones — paa46@sfu.ca
Ariel Tyson — ajt11@sfu.ca

1. Project Summary

We are building a three-class student-success predictor (Dropout / Enrolled / Graduate) from enrollment-time attributes (e.g., admission grade, attendance type, prior failures, financial-aid indicators).
The system compares two course methods we implement ourselves, custom k-Nearest Neighbors (kNN) and a custom Decision Tree and provides both local (neighbor exemplars) and global (rules) explanations. For Milestone 1, we completed the preprocessing pipeline and a working custom kNN baseline.

2. What Has Been Accomplished So Far

• Dataset: UCI "Predict Students' Dropout and Academic Success" loaded; initial EDA completed (class distribution, feature types).
• Preprocessing: categorical encoding (one-hot/ordinal as appropriate); numeric scaling (standardization); missing-value audit & strategy; stratified train/validation/test split established.
• Baseline model: custom brute-force kNN (Euclidean; tunable k) implemented and evaluated on the 3-class target; baseline metrics produced (macro-F1, confusion matrix).
• Explainability: neighbor-based exemplars demonstrated for several validation/test cases (top-k similar students and their outcomes).
• Engineering & reproducibility: repository initialized; README and requirements file drafted; plotting utilities scaffolded; nested cross-validation harness scaffolded for M2.

3. What Has Fallen Behind

Nothing material is behind: all Milestone 1 tasks have been completed as planned. Items scheduled for Milestone 2 remain on deck (Decision Tree implementation + pruning, k-d tree acceleration for kNN, full nested CV runs, calibration and cost-sensitive thresholds, and ablations).

4. Any Project Changes

Scope extension from the initial proposal: We broadened from a single-model baseline to a two-technique comparison (custom kNN + custom Decision Tree), upgraded evaluation to include nested cross-validation, per-class PR-AUC, and probability calibration (Brier score, reliability curves), and added cost-sensitive thresholds for the minority Dropout class. We also committed to k-d tree acceleration and explicit local/global explainability. These changes respond to feedback and align the project with the course's 4-person difficulty expectations.