

STAT4011 Part 1

Introduction

This project involves the classification of accounts as either *good* or *bad*, based on provided datasets. Students will apply analytical and coding skills to develop predictive models and evaluate their performance.

Project Overview

Students must form groups of 4 to 6 members to complete the project. Upon group formation, each team will be assigned a group number and a corresponding group code.

Submission Requirements

Each group is required to submit their [predictions for the test accounts](#) by 23:59 on 8 October 2025. Submissions must include the correct group code to be considered valid.

After each submission, the system will return an F1-score, which reflects the accuracy of the predictions. Groups may submit multiple entries prior to the deadline.

[A score leaderboard](#) displaying the top ten scores will be accessible throughout the project period.

Collaboration Expectations

All group members are expected to contribute meaningfully to all aspects of the project, including both coding and data analysis tasks.

Datasets Provided

- **train_acc.csv:** Contains training account data with corresponding account types. The flag column indicates classification (0 = good, 1 = bad).
- **test_acc_predict.csv:** Contains test account data. Groups must predict the flag values for these accounts.
- **transactions.csv:** Contains transaction data related to both training and testing accounts.

Grading Scheme

Final grades will be assigned based on the F1-score achieved by each group:

- Top 25% of scores: Grade A
- Second 25%: Grade A-
- Third 25%: Grade B+
- Bottom 25%: Grade B

Consultation

Every Tuesday 4:30 pm – 8:00 pm in LSB122 or on ZOOM

<https://cuhk.zoom.us/j/98686171215?pwd=LULFhZMuQ5UYoKYSmQuuyPkt6QE6Xh.1>

Meeting ID: 986 8617 1215

Passcode: 172048