Final Project

Objectives

- Apply what you have learned to analyze real data
- Interpret the results of data analysis
- Present your work

Problem

- Two datasets: *training.xlsx* ($n = 519$) and *test.xlsx* ($n = 50$)
- Variables:
  - 30 predictors (X1, X2,…,X30), all of which are numerical.
  - 1 response (Y), a qualitative variable with two classes (0 and 1).
- Test data: true values of response variable are missing.
- Apply different methods to the training data and predict on the test data using your chosen method.
- Submit a report and the predictions.

Requirements

- **Only** use methods covered in this course.
- This is a group assignment. Each group should have two or three students. Only one submission is needed for each group. Make sure names of all members are given in the title page.
- Each group is required to do independent data analysis and report writing.
- You can ask TA for help ONLY if you have difficulty in downloading a dataset and importing the data to software.
- Grading on project will be based on the quality of data analysis (70%) and accuracy of prediction (30%).

To Submit

- 1. Report: Power Point slides (pdf file)
  - Title page: course, project title, names, student ID, date
  - Page 1: team member contribution (−3 points if no)
  - Given information, problem formulation, strategy/methods, justification, data analysis, results, conclusion&discussion
  - Page limit: 10 slides (including title page); extra slides will be ignored.
  - DON'T include codes in the slides.

- 2. Appendix (pdf file): codes used in your data analysis

- 3. Predictions (excel file)
  - Put your predictions in the test data file.

Submission Timeline

- **Start time:** November 20, Monday @ 8:00 AM
  - Data will be available in *Canvas* → *Files* → *Project*

- **Submission deadline:** December 3, Sunday @10:00 PM
  - Please submit to *Canvas* → *Assignments*
    - "Project_report": submit ppt slides
    - "Project_predictions": submit excel file
    - "Project_appendix": submit appendix file