RMSC 4002 Financial Data Analytics with Machine Learning
Guidelines for course group project 2021-22 Term 1 (Fall)

The purpose of this group project is to provide a hand-on experience for students to try the techniques learned from this course on real dataset. Students are required to choose exactly **two** of the following topics:

1. Multivariate normal distribution and simulations
2. Estimating volatilities, correlations, financial time series, and Kelly's formula
3. Value-at-Risk, Risk Measures and Extreme Value Theory
4. Principal Component Analysis, Factor Models and Recommender Systems
5. Binary Logistic Regression and Multinomial Logit for categorical variables
6. Classification / Decision and Regression Trees
7. Nonlinear Classifiers: Bayes Classifiers and Comonotone-Independence Classifiers (CIBer's)
8. Linear Classifiers: Support Vector Machine (SVM)

There are several remarks on this project:

- Students need to demonstrate how the selected techniques are used to solve a real-life problem. The data has to be a real dataset of at least 5,000 records. You may use any dataset download from the internet. (See e.g. http://finance.yahoo.com/, http://archive.ics.uci.edu/ml/datasets.html or http://mlearn.ics.uci.edu/MLOther.html, https://github.com/, https://www.kaggle.com/).
- You may use the same dataset or two separate datasets to demonstrate the selected techniques.
- The final report contains introduction to your problem of interest and description of your dataset; methods to use; findings and conclusion.
- Students may use R and/or Python in their analysis.
- Students are expected to hand in their group report. The report should be around **20** pages of A4 size (tables and figures are exempted from this 20-page limit). Please also indicate the work (in %) involved by each member in the group; otherwise we assume the evenly spread of efforts among members.
- Students can form their own group but the group size should be 4 or 5 students.
- Choose **one** member in your group to submit a soft copy of your report, dataset and related R or Python programs and/or csv files and also hand in a **hard copy** of your group report.

This project consists of **35%** of your total mark in this course and to be submitted by sending us a soft copy on or before **December 10th**.

Honesty in Academic Work

As one of the primary aims of university education is to develop the ability to think independently, students must never use the work or ideas of others as their own. Plagiarism, undeclared multiple submissions, cheating or violations of rules in examinations, employing or using services provided by a third party to undertake one's work, impersonation fraud in tests and examination, and other acts of academic dishonesty are punishable under University regulations.
http://rgsntl.rgs.cuhk.edu.hk/aqs_prd_applx/Public/Handbook/document.aspx?id=1550&tv=F&lang=en
Attention is drawn to University policy and regulations on honesty in academic work, and to the disciplinary guidelines and procedures applicable to breaches of such policy and regulations. Details may be found at **http://www.cuhk.edu.hk/policy/academichonesty/.**