**Final Group Project**

Please choose **one** from the following two data sets:

1. wine-quality-white-and-red.csv. The goal is to predict the wine quality. A detailed description of the data is in the UCI website.

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

In this website, you can find some reference paper, where the authors used some ML methods. It is possible for you to use something similar. You can also try some other methods and compare with their results.

1. Data analysis jobs. The goal is to predict salary range for different types of jobs. The data are available from the following website

<https://www.kaggle.com/intelai/data-analysis-jobs>

Please use the machine learning methods to address the problem and justify your models/ methods (e.g., compare different methods).

Format: Report should be submitted in a pdf file by **Dec 4 at 9pm via Canvas**. Report that is late receives 20% off within 24 hours late and 0 after that. The report has no more than 8 pages in normal format with font size 12. All the statistical analysis should be done in R. The R programs used for data analysis should be turned in together with the report (programs not counted into the 8 page limit).

Since this is a group project, I have assigned groups on Canvas. Ideally, all group members should contribute equally in the project. In general, I will give the same grade to all group members. When submitting the report, please upload the file via Canvas. Once any of your teammates uploads the file via his/her account, the same file should automatically appear in your teammate’s account. Please test the submission option before the deadline in case there are some technical problems. The deadline is a hard deadline and there will be no extension. If you encounter some technical problems in submission, you can email your report to me but it must be before the deadline. Please make sure the names of the all group members and the NetID are listed on the first page of the report.

Note on the peer evaluation

If you feel some of your teammates make very little contribution to the project, you may send me a short peer evaluation by email after you submit the project (such as what part of the project is done by you and your teammates). I will look at the peer evaluations from all students in your group and may give different grades depending on the situations. If you feel that the contribution is roughly equal, you don’t need to send the peer evaluation to me. In this case, I will give the same grade.

Descriptive Stats

Modelling

* Multiple linear regression
* Best subset, forward selection, backward selection
* LDA, QDA, Naïve Bayes
* Lasso
* Ridge
* Principal Component regression
* KNN
* Tree, bagging, random forest, boosting
* \*\*SVM, NET

Validation: 5-Fold CV