# Executive Summary

*Jiaxi Wu*

In this assignment, we build models to forecast the credit default of public companies. To achieve this goal, we use 10 years of data about 9 financial metrics for 10000 companies to build the model. So we have at most 90 predictors and 10000 samples. Denote default as 1, nodefault as 0, we can use OLS and Regression Tree to predict a company is default or not. For OLS, we choose the last value of each financial metric as predictors, while for Regression tree, we use all the suitable variables. When assessing the performance of prediction, it's reasonable to consider MSE and $R^2$ as a measure. It's also necessary to separate a hold-out(testing) sample to assess the performance since good in-sample performance may overstate the overall performance and overfit the model. When training model, we use cross validation to choose the complexity parameters such as the depth of a regression tree.

As a result, OLS has in-sample and hold-out $R^2$ around 0.1 and MSE around 0.09. While after tuning, the Regression Tree has optimal hold-out $R^2$ of 0.2 and MSE 0.079, with tree depth 8 and minimum units in each non-terminal node 20. In-sample $R^2$ for Regression Tree is 0.62 and MSE is 0.037. Hence Regression Tree has a better performance than OLS on this data set.