**Apply machine learning to my capstone project**

Jihye Ko

The question that I have is what can be a good predictor for employee’s mental health issue in the company. To make a question more specifically, I had to find out which variables are strongly correlated. Previously, a dependent variable was determined as ‘seek\_help’ with following independent variables, like ‘wellness\_program’ and ‘no\_employees’. With those variables, I had to make a proper fit model for the prediction. To do that, I need to determine whether my problem is supervised or unsupervised. If it is supervised, is it a regression or a classification? Since my problem is to predict the target, so it is supervised learning. Also, my problem falls into classification. Therefore, tret-based models should be used for this project. The library, ‘rpart’, is a recursive partitioning for classification, regression and survival trees. In other words, classification and regression tree can be generated by the rpart package. To grow tree, I had to choose which independent variables should be used. Since I chose ‘seek\_help’ as a dependent variable, I chose independent variables based on correlation coefficients generated previously. After I made several models by adding different variables, I chose the most proper one based on their cp values. Then, the accuracy of the model was calculated by applying the model to test data. The limitation of my project is that there is no test data that I can compare. So, after a discussion with mentor, the original data was used again to get the accuracy of the predicted model that I built.