#### **Analytical Project Assignment**

Team H

Modeling quality of suggestions

## Associated data sets

suggestions.csv

# Data description

This dataset is a ‘scrape’ from an online forum of a large human resource company. The purpose of the forum is to provide a way for employees to give suggestions to the upper management about a variety of topics. Each suggestion corresponds to a particular thread on the forum (designated by suggestion ID). Other posters can respond to, vote on, or simply view the suggestions in the forum. There is also information about the author of the suggestion such as how many posts he/she has on the site and how long he/she has been an employee (in terms of the number of days with the company).

A group of interns have painstakingly gone through the suggestions and highlighted the ones they would recommend to follow up on (about 500 of them). However, this process is not scalable or sustainable. They have approached your team to develop a method that will narrow the scope of their future search so they do not have to sort through the ‘bad’ suggestions.

### Your tasks

1. Determine which combination of attributes of the suggestion (and maybe the person who wrote it) can be used to predict a ‘good’ suggestion. Does number of views matter more or less than votes?
2. How much does the ‘age’ of the employee matter when it comes to their ability to make a good suggestion? Are the employees with longer tenures making better suggestions than those with shorter ones?
3. Can the same data be used to rank employees based on their demonstrated ability to make predominantly good suggestions? Can it be used to identify groups of employees whose suggestions could be aggregated to provide more reliable suggestions than made by the best individuals?
4. Make recommendations to your IT department about better ways they could collect this data in the future. What other attributes would prove useful and why? Would it be possible to build a completely automated suggestion ranking system?