Final Project Proposal – Sako Kassabian

# There exists a problem

Students are constantly being tested. Papers continually being graded. Events being organized. Go, go, go. It’s like a machine. Every year, over and over. For the administrator, they see hundreds to thousands of students each year. It becomes a blur. Often feeling powerless in their altruistic desire to help these students overcome obstacles. It’s difficult and rare that administration would dive a bit deeper into reviewing test scores year over year and look at student growth or decline.

# How can it bring value?

Most schools are using some sort of enterprise system that tracks their student’s data such as demographics, attendance, grades, etc. Most of the time data is only travelling in one direction and is rarely looked at. Providing a way that allows schools to enter their student’s testing/attendance/ grades data in and show them information that may lead to action would be invaluable to them. By providing a year end report to schools that breaks down their scores and provides some predictive modeling, they can get in front of the students that may start to fall behind and, instead of being reactive, be proactive in their actions for improvement.

Some examples would be:

* Exploratory analysis on absent rates by zip code
* Trying to predict absence rates by test/grade scores
* Does English learner status have any effect on test scores or absence rates?
* Using a mixed effect model, can we look at the differences between test scores both at the level of the individual YOY and between students YOY (year over year)

# Where is the data coming from?

The data is being provided from an old colleague (permission obtained) that contains anonymous student data from a few schools. In production, it would be your own school. The data contains the following information: state student number, school student number, grade level, English learner status, race/ethnicity, zip, English test score 17/18, English test score 16/17, math test score 17/18, math test score 16/17, days absent, attendance rate, absent rate, semester 1 English grade 18/19, semester 1 math grade 18/19

# Techniques from the course

I plan to use data visualization, data cleaning, feature engineering, regression and classification models, and mixed effect model as my specialization. (I don’t know where else to put this but if I fail the proposal, could you please describe what you are looking for in this section other than what I have written?)

# Challenge?

Our data might be biased in demographic as some schools are heavily biased to one demographic. While this may not be a problem when taken into context, the conclusions that may arise must take this into account. We may also have a problem that each schools does not have tons of data. Some of these problems may be solved by combining other school’s data in order to supplement. I can also use models that don’t require tons of data in order to make a prediction. Therefore, models such as a neural network are not a good fit.