EECS 349: Project Proposal

Team Name: kbbz
Group Members: Kevin Chen, Basil Huang, Brittany Lee

Task

Facebook is one of the core ways people express themselves and share their lives. College students often post about their life events such as getting an internship or graduating, but on a day to day basis students often struggle to decide whether their thoughts are worth sharing. Will my friends "like" this status? Will they think it is annoying? Will they laugh? College students' social lives are measured by the judgement of others, and one simple scale is the number of likes that a status receives. Craving approval from their friends, a student may want to predict how popular a particular post would be in order to decide if it was worth sharing. Instead of impatiently waiting for the first few "likes" to appear on their status, the student could rest assured that their status would be popular among their friends.

The specific task we are going to address is predicting the popularity of a Facebook status. The popularity will be measured by the proportion of your friends that will be predicted to like your status. We will use data from the Facebook API and look at features such as the attributes of the poster and the content of the status. The initial approach will be to use the nearest neighbor algorithm.

Data

Using the Facebook API, we will access our friends lists to get the user ids of our friends. With the user ids, we will be able to access the statuses, number of likes, and other information related to the person who posted the status. The Facebook API and FQL (Facebook Query Language) can be used to generate this information dynamically, and we will store all of this information in a database.

Features

Our analysis of Facebook status popularity will factor in characteristics of both the poster and the status itself. Namely, we want to consider the number of friends the poster has, their age, gender, location, and frequency of statuses, the time the status was made, and the actual contents of the status.

Initial Approach

After gathering the data, we will handle missing values and then parse through the content of the statuses. Once this is done, we will use the nearest neighbor algorithm to group similar statuses and situations together and relate that to the number of likes. With this prediction model, we will cross validate it. We will evaluate success primarily on validation performance in addition to a few live test cases.