

Final Project - Quora Questions Pairs

- Miles Curry ([currymi](#)), Christian Armatas ([armatasc](#))
- CS434
- Dr. Xiaoli Fern
- Spring 2017

Project Proposal

Our team has chosen to tackle the Quora Question Pairs problem. At our first glance at the data we plan on using a Naive Bayes classifier on the data to determine if they are similar choices or not.

At the first look at the data we are sure there will have to be data cleaning of some kind. We assume that we will have to pick out keywords that help us best characterize the data, but as of now, we are not sure how that is. Possibly removing words that are not necessary, like articles of speech (the, and, etc.).

We are thinking that the more two questions have similar words the percentage of their class being linked would be greater. We think that using the Naive Bayes, or another form of Naive Bayes, will help us make predictions.

Our plan is to implement the Naive Bayes classifier, train and it and see if we can fine tune the data before we train and test it to get better results.