Summary of the Proposal
We will use the twitter API and a natural language processing algorithm trained on a large dataset of queried tweets to create a model that predicts the sentiment of future tweets on a scale of positive, neutral, or negative with the overall goal of improving the success rate of previous attempts.

Background
With every successive generation, more and more young people are using social media:

The amount of user-generated data being made by each generation creates both a unique opportunity and a unique problem for data scientists. What is the best way to parse, clean, and interpret this enormous mountain of data?

One way to approach this problem is through a subset of machine learning known as NLP (Natural Language Processing) which utilizes Naive Bayes algorithms and/or Decision Tree algorithms to create predictive models for processing natural language text into meaningful insight. A fantastic
data set for use in this regard is the social media platform Twitter. Twitter is a site that allows users to post small, 140 character blurbs for public viewing. These can be attached to specific, usually associated “hashtags” that classify tweets into general themes. The Twitter API allows for querying of tweets using hashtags which makes for an excellent, data-base like query system for the purpose of machine learning model training.

What we propose is to use the Twitter API and a Naive Bayes NLP Algorithm to train a model to predict the sentiment of a tweet as being either positive, neutral, or negative. Previous attempts have had varying degrees of success, and our hypothesis is that if we use more focused hashtag queries, we can increase the success rate from these previous attempts.

Goals and Objectives
The goal of this research is to use sentiment analysis on tweets so future entrepreneurs, politicians or anyone who wants to promote their personal platform can enhance their presence in the social media medium to gain the user’s trust. Specifically, we will be using a more meaningful and diverse dataset queries than what has been previously been recorded to exceed the success rate on what the sentiment predictions will be. With these datasets, we are able to distinguish the positive, negative and neutral criticisms received so the outcome will have a better turnout.

Data and Methods
- Data Set
  - Queried Tweets Data Set using Hashtag querying with the Twitter API
- Methods
  - Naive Bayes Algorithm for Natural Language Process (NLP)
- Improvements
  - Use focused queried hashtags to improve the success rate of sentiment prediction.

References