**Introduction/Problem Description**

With the rise of social media and the internet, the opinions of thousands can be accessed with just a keystroke. Never have the opinions of multitudes of people concerning a platitude of topics been so readily available. Websites such as Facebook, Twitter, and Yelp, offer platforms where people can voice their opinions and express their approval or disapproval for an issue or product. This information has been useful for the those connected to the internet and many would relate to trying to pick a good restaurant for dinner or finding the next movie to watch using these platforms as a means to access information.

However, consumers are not the only ones taking advantage of this new abundance of data. Businesses have been using these services as well to retrieve valuable feedback about their products or services. In turn, these businesses have used this feedback to either enhance their product, or to find the niche where their product will be successful. A challenge many of these businesses have run into is the sheer abundance of data and how to consolidate this data to understand the general consensus among consumers towards the product or service.

It would be unreasonable to physically sieve through each and every comment or post and we hope to solve this problem using machine learning algorithms. The goal of this project is to develop an application to process a comment and analyze the author’s opinions and sentiments based solely on word choice and sentence structure. Using the Naïve Bayes and J48 tree models the provided application will be able to determine whether a provided comment is negative or positive. Based on the accuracy of each model we will determine which is best suited for use in sentiment analysis of text.

Previously, most sentiment prediction problems have been addressed using Naïve Bayes classifiers and SVM (support vector machines). J48 trees have been used for sentiment analyses on many platforms including Facebook and Twitter. In these studies, researches were primarily categorizing sentiment data in two categories: positive/negative. For our project, we hope to use Yelp restaurant reviews to better determine the emotional state of the user. Since Yelp utilizes a 5-star rating system for its clients, our goal is to more specifically determine the author’s emotional state by increasing the number of categories.

**Description of the Data Used in the Project**

Description of the data used in the project

What have you done so far

What remains to be done