



ALY 6080: Integrated Experiential Learning

Individual Project Proposal

Submitted to

Bryce Allen

Submitted by

Md Tajrianul Islam

Date: Oct 18, 2020

## Individual Project Proposal

The blast of Web 2.0 has prompted expanded action in Podcasting, Blogging, Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. Twitter is one of the most popular microblog platforms on which users can publish their thoughts and opinions. Subsequently there has been a sudden increase of enthusiasm for individuals to mine these tremendous assets of information for suppositions. Simply put, the objective of sentiment analysis is to categorize the sentiment of public opinions by sorting them into positive, neutral, and negative. By utilizing the sentiment analyses the user can able to know the criticism about the services or item before buying it and the firm can know about the feeling of clients about their items, with the goal that they can analyze consumer satisfaction and according to that they can improvise their items. Sentiment analysis solves a number of genuine business problems, such as:

- It helps to predict customer behavior for a particular product
- It can help to test the adaptability of a product
- Automates the task of customer preference reports
- It can easily automate the process of determining how well did a campaign run by analyzing the sentiments behind the reviews from a number of platforms

Our goal will be to create a prediction algorithm that can help a brand better understand their social media presence/ reviews/ focus group reports through classification. I am going to perform several steps of text pre-processing and then experiment with multiple classification mechanisms. Using a dataset of 1.5 million tweets and TFIDF features, we will be comparing the accuracy obtained using various classifiers for this task.

As the data set is quite large we are planning to use R and Spark for the project. We will be using Sparlyr which is an interface that connects R with Spark. It filters and aggregates Spark datasets

## Individual Project Proposal

then bring them into R for analysis and visualization. Although there are some natural language processing libraries available in python like Scikit-Learn library, which makes the job of building the classifier easier. As there are also options to work with Python within Spark, it would be very obnoxious to leave out the probability of using python as we progress with the project.

Dataset URL: <https://www.kaggle.com/kazanova/sentiment140>