Arteen Zahiri, Jenny Son, Jinna Oh

Professor Zhi Li

MIS3640-01

9 October 2019

Term Project Proposal

1. **The Big Idea:**

The main objective of our project is to provide a simple and straightforward website that can provide sentiment analysis for any keyword users put in. Sentiment analysis is a process of analyzing the emotional tone behind the series of words. Through sentiment analysis, users can identify if the keyword is positive, neutral, or negative. Moreover, it is one of the most commonly used tools for analyzing opinions about a specific topic. We wanted to create an accessible and easy-to-use website that does the hard work for numerous companies that conduct sentiment analysis. Rather than having to create a whole programming application to conduct sentiment analysis, companies can simply go on our website to get sentiment-related information about any topics.

The sentiment analysis would be based on tweets posted on the Twitter server. Since Twitter is one of the most popular social media platforms, nowadays, we believe that Twitter would provide a holistic, comprehensive view of the overall sentiment of the public. We will provide three specific analyses, which includes polarity, subjectivity, and the overall sentiment. Our minimum viable product would be a website that provides the basic data regarding sentiment analysis such as polarity and subjectivity. For our stretch goal, we plan to provide visualized data analysis of the sentiment, which would include pie charts and scatter plots that highlight historical changes in the sentiment of a specific keyword. At that point, we may have to consider creating a business and charging political campaigns and think tanks to access our technology.

1. **Learning Goals:**

In today’s digital world, social media is an integral part of life; and as politics become highly involved in utilizing social media to sway votes and releasing content, it is crucial to analyze how such content is affecting citizens. By doing so, organizations and people can gauge how the public is reacting to news and events. This type of analysis is crucial for political campaigning and strategic policy framing. Outside of politics, businesses and organizations massively crave this data to perceive public sentiment and identify successful marketing strategies. Thus, the learning goals are to utilize python in order to gather and analyze user data through Twitter. Here, we must specifically learn to collect the right data: user sentiment. The team will learn to scrape and filter the right terms that display a user’s emotions. Ultimately, each of our members will understand the principles of sentiment analysis and specifically how to use python to conduct it and use it to make predictions, solve problems, and make strategic decisions.

Further, data is never complete without visualization. We believe creating a website that displays the user’s sentiment through a website is a difficult learning experience. This means we must learn to automate the website like how python would run the analysis. Due to the large selection of words and data on the web, automating this process will be significant. In effect, we are taking large quantities of raw, unorganized data and converting and visualizing it into actionable insights. This kind of skill set is crucial in today’s business world and can be applied in a vast variety of industries and conditions. While each of us may seek different career paths, we can all link a multitude of ways this type of analysis will give us a notable advantage in our future jobs.

1. **Implementation Plan:**

Potential implementation plan:

* Import tweepy and textblob
* Create a developer account for Twitter to be able to scrape tweets
* Write a script that provides sentiment analysis for a specific keyword
  + Possibly utilize the script we use for our future class: “Data Analysis Using Python”
* Create a website where users can input keywords that they would like to analyze
* Integrate the python script to the website that we have created
* Find a way to automate the running of the python script that does the analysis.

1. **Project Schedule:**

**Week 1:** The first week we will create developer accounts for Twitter in order to scrape online Tweets. Then we will import tweepy and textblob.

**Weeks 2 & 3:** We will be writing a script that will provide sentiment analysis for specific keywords.

**Week 4 & 5:** Build a website that integrates the python script for the website. The website will display sentiment analysis for any user input of a keyword.

**Week 6:** We will look for bugs to ensure our code runs seamlessly.

**Week 7:** User testing

**Week 8:** Presentation Preparation

1. **Collaboration Plan:**

We plan to follow a hybrid collaboration approach so that tasks can be completed on time and effectively. As we break down the project process into steps, we will decipher which steps are best worked on as a team and which can be split and completed independently. In addition, we can also determine tasks that can be done individually and in parallel. Since we have a multi-faceted project, this seems like the most efficient avenue.

Our project has three key components that consist of multiple parts which we must accomplish: 1) use python to code a script for Twitter sentiment analysis 2) create a website 3) integrate the python application into the website. The first component can be divided up into several tasks which we can do individually. Some of these tasks can even be done simultaneously. The second and third components will require more of a group effort. For the second, developing a website is a multi-layered process that will mandate group collaboration for design. For the third component, this is the most challenging piece of the puzzle and in fact, we may even need to recruit the services of the esteemed Professor Li himself to assist us in understanding how to use Flask or some other web framework for python to complete the integration process.

As for asking a future employer, we will ask a Tinder software engineer to guide us through making this a concise yet efficient code. We believe asking such an engineer will help us during collaboration, by asking us the right questions and advising us through the vast amount of relevant data. On a daily basis, Tinder is analyzing billions of data, also utilizing sentiment analysis to create better matches for their users. We believe this collaboration will be useful.

1. **Risks:**

The biggest risk to the success of this project lies within integrating the python code into the website and ensuring a successful web application. None of our team members has experience in this field and it is the most important part of the solution, as without it the solution is not dynamically accessible for public use. If learning how to use a python web framework proves to be more challenging than expected, we may be in for a rough ride.

Another risk we may face is the process of filtering the data. Many emotional keywords could be analyzed incorrectly and out of context. For example, “I hate how happy you make me feel”. In this context, hate should be counted as a positive term, rather than negative. It remains to be seen if our code is robust enough to correctly analyze such example. Similarly, we feel this type of issue could arise when detecting sarcasm in tweets, and whether or not our code can consistently recognize if the statement is meant to be taken positively or negatively.

1. **Additional Course Content:**

We think it would be beneficial to the class and important to cover python web frameworks, such as Flask, Django, Pyramid, etc. Obviously this would be especially helpful for our project specifically, but we think it is an extremely useful tool for everyone taking the class. Frameworks are great for developers as they offer a structure for application development. They automate the implementation of common solutions, cutting development time and allowing developers to focus more on application logic instead of routine elements.

We believe one of our future classes that does data analysis using python would be extremely helpful for our project. Learning how to analyze data using python would allow us to write a coherent script that provides sentiment analysis of the keywords that users plan to analyze.