**Twitter Sentiment Analysis of Climate Change**

**Project Update:**

For this iteration of my project, I extracted my dataset from Twitter using my API credentials. My project is slightly more complex because I do not have a pre-made dataset ready to use, I have to extract it. In order to do this, I had to implement various different techniques and using different methods. Some problems that I have encountered during this iteration of my project is that I severely underestimated the number of tweets being made a day regarding climate change. This has thrown me for a loop, and I had to reconsider my original thought process of extracting tweets dating back 5 years. I have chosen to limit the tweets I analyze going back exactly a year. This limits the number of tweets and also I can analyze tweets that are relatively new.

**Literature:**

There are various different projects that have conducted Twitter sentiment analysis, using various different techniques to classify their tweets. When analyzing the projects online, I learned that many of the projects used Natural Language Processing [1] for the analysis of tweets, which I plan to use as well.

One project did not only classify the tweet but calculated the sentiment of each tweet and built dashboards to visualize it [3]

The difference between the projects I have researched and mine is that I plan to use not only Natural Language Processing for creating my classification, but also Naïve Bayes. This is so I can experiment with these two popular classifications.

**Data Cleaning:**

The hardest part of the data cleaning was the extracting of tweets. I had to find a way to extract the tweets and import them into a CSV. The way that I found worked the best was using a package called Tweepy. Once I extracted the tweets with the variables I want, such as location of the tweets and date of the tweets, I exported that into a data frame for further analysis. Additionally, I exported the tweets into a CSV for data storage.

**Next Steps:**

Conduct further data analysis of my dataset and improve the way I am listening to my tweets. In my next step, I plan to start my Naïve Bayes classification program.

**Link to presentation:**

**References:**

**[1]** [**https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/**](https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/)

**[2]** [**https://towardsdatascience.com/twitter-sentiment-analysis-classification-using-nltk-python-fa912578614c**](https://towardsdatascience.com/twitter-sentiment-analysis-classification-using-nltk-python-fa912578614c)

**[3]:** <https://realpython.com/twitter-sentiment-python-docker-elasticsearch-kibana/>