**Determining the Political Alignment of a Person Based on Twitter Data**

Adrian Vasu

**Background:**

Twitter was founded on March 21, 2006 and the first tweet was also sent on this date. Since then in the last 14 years twitter has become a daily habit for more than 145 million people across the world and 330 million monthly users worldwide. Twitter is a social media platform that allows for the sharing of small 280-character messages (sometimes including pictures) to a network of followers which you accumulate over time. These messages are known as “tweets” and are sent out instantaneously. While twitter has a set of guidelines for its users to follow regarding what they can post recently twitter has become a political hotbed with many people getting into heated discussions about controversial topics. This recent influx of political material into the platform is also motivated by many politicians taking to the platform to share their messages with the “world”.

Python 3.9 is a programming language which many scientists and programmers take advantage of because of the various useful tools which python allows one to utilize in various projects. One module which is particularly important especially in the field of data science is SciPy. SciPy has various machine learning methods built in that one can use to analyze and report trends in data sets. Taking advantage of the SciPy toolkit also allows for the creation of graphics which represent the trends in the data. Overall making use of the python SciPy toolkit in the implementation of the project should allow for the efficient application of machine learning tools to determine a person’s political alignment based on their twitter data.

**Motivation:**

With the recent influx of all the political messages into twitter I am seeing some great discrepancies between what a person is stating their alignment is and the material which they are posting and supporting. This discrepancy should have a way to be quantified and I would like to know how accurate someone can predict what their political preference appears to be to someone looking from the outside in. This combined with my general interest in politics and programming/computer science makes a machine learning implementation the logical next step in my data science education.

**Scope and Objectives:**

To tackle this project I would like to narrow down the scope to ensure that I have chances to finish it in the semester. The main objective will be using the Twitter API for python to pull in and clean data and then feeding it to a machine learning algorithm within SciPy to build a basis that determines a person’s political affiliation based on their tweets. The main objectives within this would be to successfully:

1. Implement the Twitter API into python and be able to download a person’s complete twitter history
2. “Clean” the twitter data to be able to parse based on a dictionary of created affiliations
   1. i.e.) Classify each tweet on a rolling scale 100% liberal to 100% conservative
3. Prepare the data in a way to feed and teach the machine learning algorithm (approximately 20% of useable data)
4. Develop the algorithm to work on the other 80% of the data and classify each person on a rolling scale from 100% liberal to 100% conservative
5. Refine the algorithm to within 20% accuracy on definitive people
   1. i.e.) George Bush is Conservative, Obama is Liberal etc.
6. Produce nice output and graphs which show the overall percentages based on a wide range of selections
   1. i.e.) All 100 senators and their listed affiliations vs. algorithm output
7. Apply the algorithm to whomever we wish ☺

**Timeline:**

1. Project Proposal – March 20, 2020
2. Twitter API – End of March
3. Partial Draft – March 30, 2020
4. Cleaned data – First week of April
5. Peer Review – April 6, 2020
6. Taught Algorithm – Third Week of April
7. Refined Algorithm – End of April
8. Output – First week of May
9. Final Project – May 5, 2020