Initial Plan:
Mapping COVID-19 sentiment on Reddit

Oliver Carter

C1835576

Supervisor: Steven Schockaert
Moderator: Amir Javed

Background

On March 23rd 2020, the British Prime Minister Boris Johnson announced that every UK resident should not leave their place of residence unless absolutely necessary, signifying the beginning of the UK's first national lockdown in a bid to combat the looming pandemic. From our early understanding of the new coronavirus, theoretically, if we incorporated a strict national lockdown thorough enough that every person who were infected received enough support to be willingly able to self-isolate until they were no longer infectious, after a few months the virus should surely have been weeded out as quickly as it came.

Project Description

This project is aimed at finding to what extent Reddit can be used to measure sentiment and compliance towards COVID-19 restrictions in the UK. I will be using reddits python api to collect submissions from subreddits across the UK which are the subject of COVID-19 restrictions. The submissions I am most interested in are the ones which are announcing new local restrictions because then I can process the sentiment of the comments to evaluate the general publics' response to tighter restrictions. It would certainly be interesting to see if there are any trends that could be uncovered in this data to increase the utility of Reddit. For example, I will be looking to see if the model can accurately predict whether a county has had restrictions increased by detecting increases in negative sentiment in certain areas across the country. I will be visualising this data on a heat-map which will show how the sentiment towards COVID-19 restrictions changed over time across the UK but also have a view to creating multiple graphs with the collected data to further increase the utility of Reddit to measure sentiment and compliance towards COVID-19 restrictions. One particular issue I find will be interesting to tackle is how to identify when a post is referring to a COVID-19 restriction. I am going to be initially looking into training a model to recognize keywords in a submission title to try and detect if it is about restrictions. However, I will also be building up a database of key events in the UK's COVID-19 timeline to narrow the search.

At present, I do not believe it will be necessary to submit an ethics approval form for this project before beginning data collection as I am not collecting and storing any personal dat. However, I will review this and check to see whether this needs to be done before data collection and if so, I will make amends in the final report.

I already have experience using python's reddit API 'PRAW' to extract information from subreddit submissions, Vader sentiment analysis, geocoding in python with GeoPy, and the natural language processing module SpaCy, which is why this particular project proposal appeals to me.

Aims

- Create an interactive map that shows sentiment towards COVID-19 restrictions across the UK
- Visualise how the sentiment towards COVID-19 restrictions changed over time
- Conclude to what extent reddit can be used to measure sentiment and compliance towards COVID-19 restrictions

Objectives

- Create a program which retrieves submissions from subreddits across the UK where the subject refers to a COVID-19 restriction
- Classify the sentiment of every post to build a picture of how positively or negatively people reacted.
- Create an interactive heat-map in flask that can be controlled to show changing sentiment towards restrictions over time
- Create graphs using collected data

Work-plan

Week 1-2 (8th - February - 19th February):
- Begin creating a prototype reddit scraper using the python reddit API *PRAW* to test collecting a sample of submissions and their comments from various subreddits.
- Test outputs from prototype reddit scraper
- Draft a layout for the final report

Week 3-4 (22nd - February - 5th March):
- Begin researching ways of creating a natural language processing model to observe patterns in submissions which announce new COVID-19 restrictions to automatically identify relevant submissions
- Implement off-the-shelf sentiment classifier to classify the sentiment of any given text
- Test classifier by passing through raw data collected from praw
- Report writing

Week 5-6 (8th - March - 19th March):
- Continue research into method to detect posts about COVID-19 restrictions
- Research methods for plotting and visualising geographical location data on a graph within python
- Attempt to train and implement natural language processing model to detect if the subject of a submission is about COVID-19 restrictions
- Report writing

Week 7-8 (22nd - March - 2nd April):
- Implement a prototype method for visualising some sample data from reddit that has been classified with sentiment analyser
- Design a prototype flask website to display the raw extracted data
- Report writing

Week 9-10 (5th - April - 16th April):
- Fully implement an interactive map into the flask website with all intended functionality
- Test final flask project and document results
- Report writing

Week 11-12 (19th - April - 30th April):
- Create graphs for the report with processed data
- Report writing

Week 13-14 (3rd - May - 14th May):
- Finishing touches to final report and proof reading

Meetings

In our initial meeting I have discussed with my project supervisor about the regularity of our meetings. We have already arranged a meeting on the 12th of February but for now I do not think we need to have regular meetings because I am confident in my plan for the project and in my ability