Full name and semester: Srikeerthana Reddy Bandi (Spring'23)

Title: "Public sentiment towards climate change on Reddit": A sentiment analysis study of posts and subreddits related to climate change. (467 words)

Research Question: How do public sentiments towards climate change differ across various subreddits, posts and comments related to climate change on Reddit?

Background: Climate change is causing changes to the Earth's weather patterns, which can result in more extreme weather events, rising sea levels, and other serious problems. People are talking about climate change and sharing their opinions and concerns about the issue. Some are using social media to raise awareness and promote action, while others are using it to express skepticism or denial. Exploratory data analysis revealed that there is lots of discussion about climate change on Reddit (Treen, K., Williams, H., O'Neill, S., & Coan, T. G., 2022).

By analyzing the posts and comments related to climate change on Reddit, this proposal aims to find out what the public thinks and feels about this issue. As more and more individuals become aware of the urgency to take action to address this challenge, it has become critical to understand public opinion and sentiment on this topic. In order to mitigate against further damage related to climate change and this can be achieved through investigating the sentiment of the population and policy changes. (Sham, N. M., & Mohamed, A. H., 2022**).** This study aims to fill this gap by conducting a sentiment analysis of posts and subreddits related to climate change on Reddit.

Data: The data for this study will be obtained by using Python programming language and its libraries, such as PRAW (Python Reddit API Wrapper), to scrape the Reddit comments posted by users. I Will collect at least 500 data points related to climate change discussions on Reddit. The collected data will include posts, comments, and metadata information. I will focus on analyzing data from relevant subreddits related to climate change, including r/ClimateChange, r/GlobalWarming, and r/Environment and from other relevant subreddits that discuss climate change topics. I intend to gather subreddits that have minimal spam activity and fewer spam users. In case there are any spam users or posts that are still present, I will remove them during the preprocessing stage. The data collection process will be conducted using Python scripts and relevant libraries to ensure data quality and consistency.

Method:
To better understand the sentiments of people on Reddit about climate change, will first clean the data by removing irrelevant words, spam data and organizing it in a structured format. The Natural Language Processing (NLP) techniques will be used to preprocess the data by removing stop words, stemming, and tokenizing the text data. I will use a tool called VADER to understand

whether comments are positive, negative, or neutral about climate change based on the tone of the content. It will give us a score between -1 and +1 to show how positive or negative the comments are. To delve deeper into the emotional content of the posts, comments, and subreddits related to climate change, I plan to utilize a Python library called EMPATH. With EMPATH, I aim to categorize the emotions into specific topics. These results can be analyzed and visualized through various plots, such as bar charts, to identify the distribution of sentiments.

References:

1. Sham, N. M., & Mohamed, A. H. (2022). Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches. Sustainability, 14(8), 4723. https://doi.org/10.3390/su14084723

2. Treen, K., Williams, H., O'Neill, S., & Coan, T. G. (2022). Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate?. Environmental Communication, 16(5), 680-698. https://doi.org/10.1080/17524032.2022.2050776