SDA250 Assignment 3

In order to get a better understanding of attitudes towards the pandemic in the United States, our group decided to conduct a sentiment analysis towards COVID-19 in America. We used a data set of approximately 10,000 tweets that were tagged with sentiment and were on the topic of COVID-19. Our sentiment analysis tagger found that the overall sentiment of our tweets was positive. When comparing this to a poll conducted by the New York Times of over 4,000 Americans, the results were comparable, however not what could be classified as the same. The poll results indicated that people felt differently about the pandemic, and a concrete positive or negative sentiment was not evident from the results. From the tweets and questions in the poll, we infer that a positive sentiment would be the feeling that the pandemic will one day end or society will return to some form of normalcy. When asked such questions, 37% of people indicated that they felt the pandemic would end, whereas 56% were uncertain. Similarly, 41% of respondents were certain that life would return back to normal, whereas 54% were uncertain. While the results of the survey appear to return more divided results than our sentiment analysis, there are a few reasons as to why this could be the case.

First of all, the survey was taken at a time when the Omicron wave was overwhelming, leading our group to think that an event such as this would persuade people to be less optimistic. Further, the tweets were taken over a course of a year meaning that tweets would have been scrapped from the summer when there was a high sense of optimism that the pandemic was drawing to a close as vaccinations were rolled out. It is also important to consider that people usually take to social media when they have strong emotions (either positive of negative), as such, it is likely that poll data could provide a more enhanced picture of what the general population is thinking who may not have incredibly strong feelings about it or be a vocal person in general. With all this considered, we can infer from the polling data and our sentiment analysis, people have a slightly positive sentiment towards COVID-19 in the sense that there is hope that the pandemic will one day end or that life will return to some degree of normalcy.

Using the classifier model that we had trained in Assignment 2, we utilized NLTK functions like punkt, wordnet, taggers, and word lemmatizers to import, separate and tag, and clean twitter sample data from NLTK. Once the tweets were tokenized and provided a positive/negative tag, the model analyzed the overall tweet labelled it one or the other.

References:

Daityari, S. (2019, September 26). How to perform sentiment analysis in Python 3 using the natural language toolkit (NLTK). DigitalOcean. https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk

How To Perform Sentiment Analysis in Python 3 Using the Natural Language Toolkit (NLTK) | DigitalOcean. (n.d.). Retrieved March 10, 2022, from https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk

Morning Consult & New York Times (2022, January 14-16). National Tracking Poll. Morning Consult. https://assets.morningconsult.com/wp-uploads/2022/01/24150429/2201068\_topline\_PARTNERSHIP\_COVID\_SURVEY\_Adults\_v1\_SH-1.pdf

Preda, G. (2021). COVID-19 All Vaccine Tweets [Data set]. Kaggle. https://www.kaggle.com/datasets/gpreda/all-covid19-vaccines-tweets

Preda, G. (2021). Covid-19 Vaccine Tweets with Sentiment Annotation [Data set]. Kaggle. <https://www.kaggle.com/datasciencetool/covid19-vaccine-tweets-with-sentiment-annotation/metadata>

SDA250 Lab 3 Code (Text Retrieve Function)

Yazanshannak and Shurrab, S. (2021). US Covid Tweets [Data set]. Kaggle. https://www.kaggle.com/datasets/yazanshannak/us-covid-tweets