

Analyzing COVID-19 Tweets

Abstract

The goal of this project was to use unsupervised learning techniques to attempt to draw some meaning from a huge corpus made of tweets regarding Covid-19. This project sparked my interest because I lived through the whole pandemic and have seen how panic in 2020 has turned into nonchalant, casual mentions of Covid-19 in 2022. I wanted to revisit the early days of the pandemic and see what people were most concerned about by checking their tweets.

Design

This project was something I wanted to do when I found a nice dataset of tweets that had already been collected. I thought it could be interesting to see what sort of things people were tweeting about with regard to covid-19.

Data

This dataset contains approx. 45k tweets from Twitter and some other columns like Location of tweet and sentiment but I focused mostly on the original tweet.

Algorithms

- Unsupervised learning model
 - Dimension Reduction, Clustering

Tools

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling/text processing
- Matplotlib for plotting

- NLTK for text processing

Communication

A five minute Google Slides presentation displaying my findings.

The project will be displayed on Github as well.