274 project proposal

Team

Hanyan Wang (hanyanw3)

Yukun Shi(yukuns10)

Yifei Li(yifeil27)

Dataset:<https://www.kaggle.com/datatattle/covid-19-nlp-text-classification>

Introduction

There are more than 100M COVID-19 cases around the world until January 2021, and many more people are suffering from it. Though several vaccines have come out, there is still a long way to go before humans conquer it.

Twitter is one of the most popular social networking sites with millions of users and posts every day on a variety of topics. The tweets can be a powerful indicator of the trend and concerns from the public. Performing sentiment analysis on the recent posts, we can find out the user’s idea and attitude during the pandemic period.

The dataset collects over 40,000 tweets during March and April last year, not only from the United States but all over the world. It divides all the tweets into 5 categories: extremely positive, positive, neutral, negative, extremely negative.

We plan to use different models to fulfill the task as well as evaluate their performance, such as accuracy, recall and precision.