Twitter Emotion Analysis – An LSTM Approach

Rohan Allen (18BCI0247), Guide: Dr. Manikandan N | SCOPE

Introduction

Many individuals throughout the globe are increasingly utilising blogs, forums, and social media sites like Twitter and Instagram to express their beliefs to the wider world. Social media has evolved into arguably one of the most powerful communication platforms known to mankind. As a consequence, a significant amount of data is created, referred to as big data, and sentiment analysis was developed to make sense of this huge amount of data.

Motivation

This project aims to delve deeper into the topic of sentiment analysis and classify a sentiment into specific emotions such as love, anger, fear, joy, etc. This would certainly help in providing a more nuanced understanding of the mood of the public towards a certain product/idea.

SCOPE of the Project

The primary goal of this project is to stream real-time COVID-19 related tweets using Twitter and its Python API-Tweepy and then conduct a sentiment and emotion analysis on the same. The results are compared with a corpus of tweets collected two years ago. Initially lexicon-based tools such as Text2Emotion, VADER, TextBlob, etc were considered. A bi-directional RNN LSTM neural network model was trained on a labelled emotion dataset to predict the emotions on our tweet collection.

Methodology

To build our LSTM Emotion detection model, first we need to pre-process the training dataset according to Fig2. Then we label encode and tokenize so that the words are converted to a format that the machine understands. Word embedding is then done using Word2Vec so that relationships between the words can be established. Then we train the Bi-directional RNN-LSTM model with the processed data and then test the data to assess its accuracy while functioning in the real world. Hyper Parameter tuning can be done to optimize the models accuracy

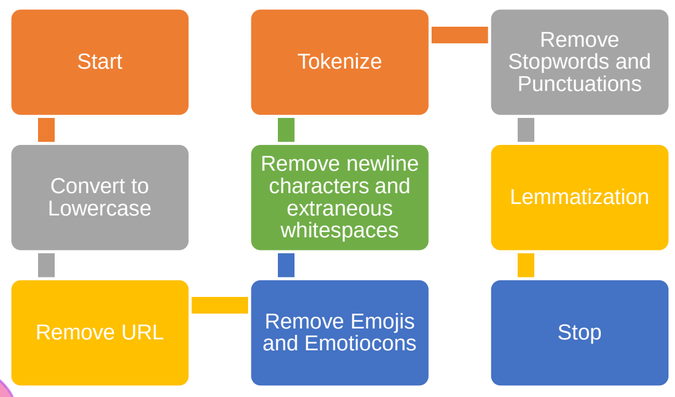

Fig 1- Pre-Processing Block

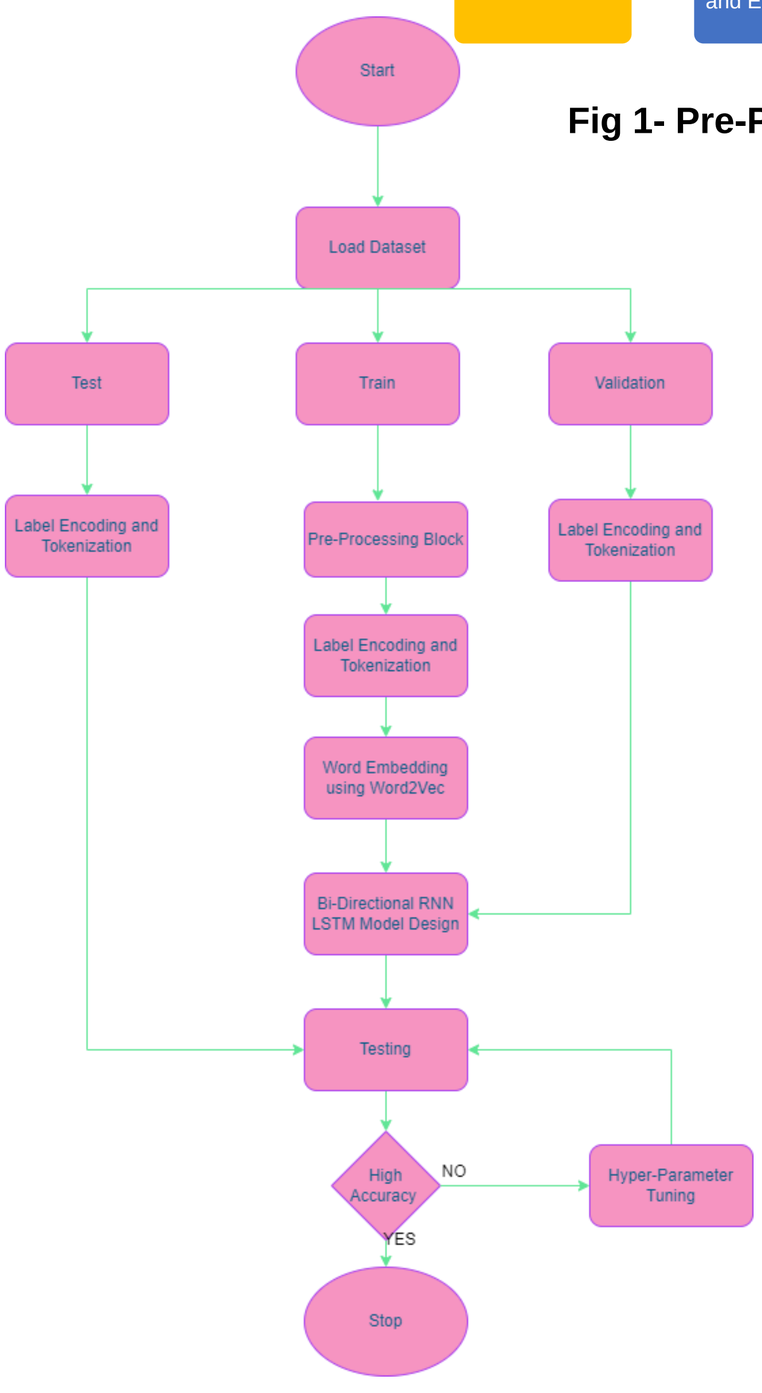

Fig 2- Flow Architecture

Results

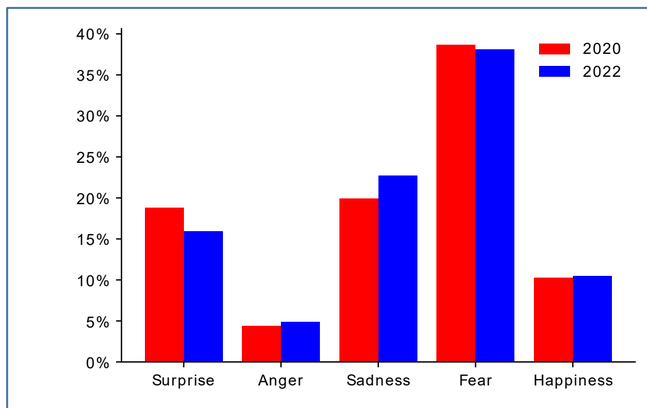

Fig 3- Text2Emotion Emotion Frequency

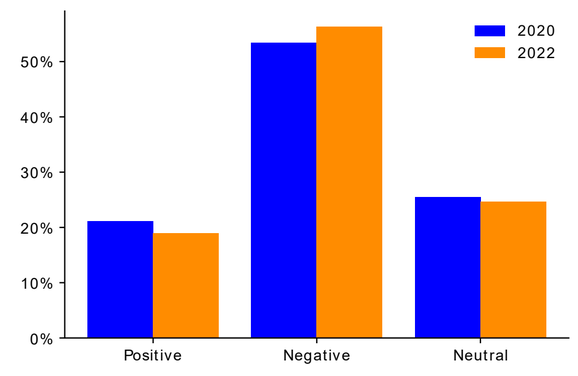

Fig 4 – Polarity Distribution

Fear followed by sadness is the most dominant emotion in both 2022 and 2020 according to Text2Emotion(Fig3). According to SentiStrength the overwhelming majority of the tweets were negative in both 2022 and 2020 (Fig4). Our LSTM model illustrates that the majority of tweets denote anger, followed by fear and sadness (Fig 5). The most prominent terms in the 2022 wordcloud weighted according to TF-IDF include covid19, longcovid, covidisnotover, omicron, vaccine, mask, etc (Fig 6).

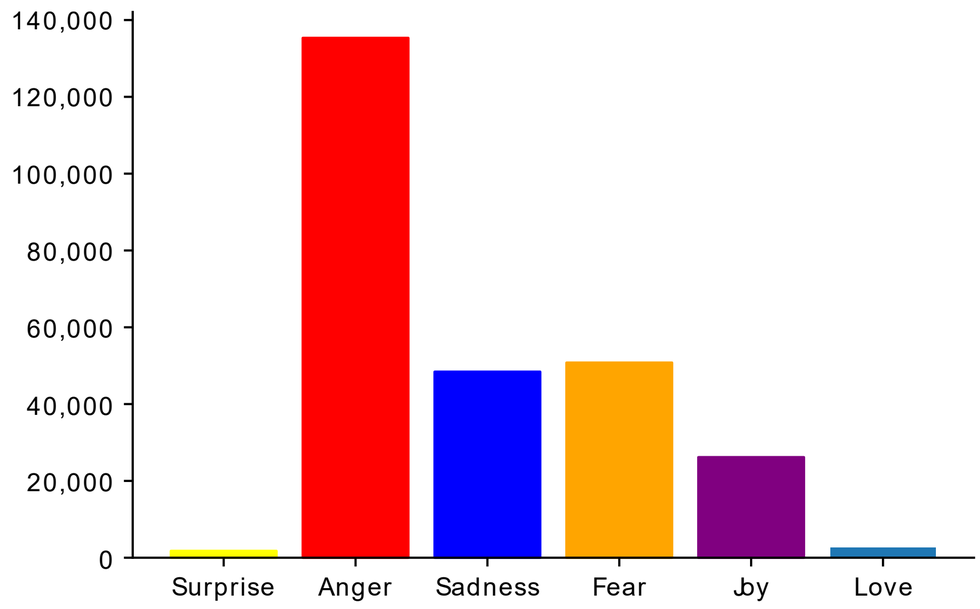

Fig 5- Histogram Depicting Emotion Frequency Distribution

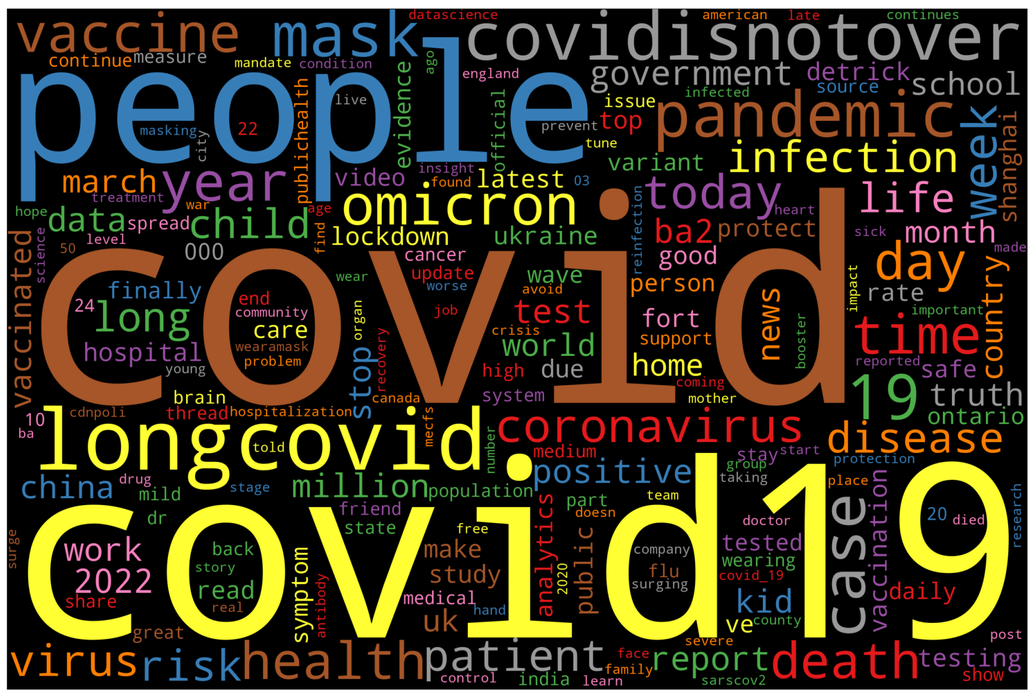

Fig 6- Word Cloud

Conclusion

We were successful in building a model that predicts the emotions of given tweets, with an exceptional accuracy score. One pertinent observation we noted was that our RNN model predicted anger to be the most dominant emotion in 2022. Compared to 2020, fear was the most dominant emotion according to Text2Emotion. This makes sense as with the passage of time, people may have gotten accustomed to and fed up of COVID-19.

References

M. A. Palomino and A. Padmanabhan Varma, "Any Publicity is Good Publicity: Positive, Negative and Neutral Tweets Can All Become Trends," 2020 39th International Conference of the Chilean Computer Science Society (SCCC), 2020, pp. 1-8, doi: 10.1109/SCCC51225.2020.9281266.

. N. Colnerič and J. Demšar, "Emotion Recognition on Twitter: Comparative Study and Training a Unison Model," in IEEE Transactions on Affective Computing, vol. 11, no. 3, pp. 433-446, 1 July-Sept. 2020, doi: 10.1109/TAFFC.2018.2807817.

Saravia, Elvis & Liu, Hsien-Chi & Huang, Yen-Hao & Wu, Junlin & Chen, Yi-Shin. (2018). CARER: Contextualized Affect Representations for Emotion Recognition. 3687-3697. 10.18653/v1/D18-1404.