Paper Review 01

Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm

Md. Alahi Almin Tansen (190321028)

Department of Computer Science and Engineering, European University of Bangladesh

CSE – 459: Internet of Things

Date: 31-03-2023

Purpose of This Research

This research is done to get the sentiments of general people towards the Pfizer, Moderna, and AstraZeneca COVID-19 vaccines. The authors collected tweets from Twitter by using an authentication token for Twitter API and then analyzed them using Natural Language Processing techniques. After processing the data, a supervised KNN classification algorithm was used to sort it into three categories: positive, negative, and neutral. The sentiment categories of positive, negative, and neutral were assigned to the processed data. The results showed the positive, negative, and neutral sentiment rate towards each of the three vaccines.

Proposed System

This paper proposes a system for analyzing tweets about the three vaccines. Authors uses Natural Language Processing techniques and KNN classifier to determine the sentiment of each tweet. They used Tweepy library for collecting data and saved in CSV file format. Preprocess includes tokenization, normalization, and lemmatization. Polarity and subjectivity are determined, and KNN classifier is used to classify the polarity data then they visualized and compared it for final result.

Architecture

There is no specific architecture used in this research. The researcher used Natural Language Processing (NLP) including text conversion to lower case, stop word removal, fixing misspelled words, replacing emojis with plain English, removing special characters/URLs/HTML tags, tokenization, normalization, and lemmatization. Object identification is also done. In KNN they load the dataset, particular K value is used. Distance is determined using Euclidean distance and sort it. After selecting top K rows, most frequent class is being observe by assigning the data point, then end the algorithm.

Experimental Procedure

Researcher used different visualization techniques to analyze the processed tweet data related to three vaccines. For visualizing the most commonly occurring words and polarity/subjectivity scores word cloud and bar diagrams are used. Scatter plots are also used for a better understanding of frequency distribution. Additionally, maximum average polarity scores for 10 tweets are calculated and visualized separately for each vaccine. Then converting the data into polarity scores. The KNN classification algorithm is used to classify the scores into three sentiment.

The resulting classification for Pfizer, Moderna, and AstraZeneca showed that people have generally less belief in AstraZeneca vaccine compared to Pfizer, Moderna vaccines.

Future Plan

The results show that Pfizer had a positive sentiment rate of 47.29%, Moderna had a positive sentiment rate of 46.16%, while AstraZeneca had a positive sentiment rate of 40.08%. Therefore, people have a higher positive sentiment towards other two compared to AstraZeneca. These findings can help authorities to provide people with the vaccine they trust, which may lead to peaceful control of the pandemic. More research can be done towards sentimental analysis to determine newly invented vaccines and current vaccine performance.

References

1. Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm by FMJM Shamrat, Sovon Chakraborty, MM Imran, Jannatun Naeem Muna, Md Masum Billah, Protiva Das, OM Rahman