Modelling Public Mood to Detect Events

PRASHANT TOMAR* and ROHIT CHAUDHARY*, Old Dominion University, USA

We investigate sentiment analysis using data from Twitter. The analysis was carried out using the Twitter corpus, which was compiled from the many timelines in accordance to the event. We are going to perform a psychometric analysis on the information that has been collected from Twitter in order to identify different mood states, and then we are going to calculate a mood vector for each timeline of the event. Positive, negative, and neutral are the three states of mind that will be utilized in the course of our investigation. These moods are described as follows: We shall make an effort to determine the effect that happenings in the social, economic, and political spheres have had on the myriad facets of the general public's disposition. After that, we are going to make an effort to conduct a large-scale analysis of sentiment, with the goal of constructing a stable platform on which to model intense patterns in terms of their predictive potential, as well as the sentiment both during and after an event.

ACM Reference Format:
Prashant Tomar and Rohit Chaudhary. 2022. Modelling Public Mood to Detect Events. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, April 30–May 6, 2022, New Orleans, LA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Twitter is a social networking website that also offers micro blogging services. Micro blogging allows users to publish their own brief posts, known as tweets, to the website. These tweets typically include text, videos, photos, and links. People utilize these services for daily chatter, dialogues, information sharing, and news sharing, which leads us to the topic of whether or not the sentiment of the public's tweets might assist determine an event that will take place in the near future. Twitter empowers its millions of users to communicate with anybody, anywhere in the world, but it also exposes them to potential dangers, such as the dissemination of fake news and disinformation, harassment and nasty remarks, data security, and privacy. It has been observed that tweets typically belong to one of two categories: either the user is talking about themselves, or the primary purpose of the tweet is to convey information. In either scenario, tweets can be understood as a microcosmic representation of the prevailing sentiment. The purpose of this study is to investigate the relationship between the current economic and political crises in Sri Lanka and the shifting attitudes of the general population, which we will determine by conducting a sentiment analysis of tweets produced between May 02, 2022 and Aug 16, 2022. On the basis of a huge Twitter corpus consisting of published tweets, we will especially investigate the interaction between the mood of the general public and the results of events.

2 RELATED WORK

In recent years, a great deal of focus has been placed on developing an understanding of human behavior through the use of the approach of sentiment analysis. Several approaches have been tried out in order to get a better grasp on event

*Both authors contributed equally to this research.

recognition and sentiment analysis. In this investigation, we took a look at a few different research publications that focused on event identification and sentiment analysis. The purpose of this section of our study is to offer a concise analysis of event detection strategies. A significant amount of study interest has been directed toward the overarching theme of extracting information about real-world occurrences from social media. The real-time detection and tracking of events, the analysis of social media, the summarization of microblogs, and the visualization of information have been the primary focuses of research efforts.

Studies that were happened in the previous years have shown Twitter's key role in recent protests such as those led to Arab Spring [4, 10], London Riots [7] and Thailand Protests [2]. Other studies have perform tweet content and sentiment analysis during events and protests[2], investigate twitter usage at time of event[9], describe, model, and interpret the user networks and relationships between social sites as well as social movements[6]. Mohsen et al. [1] perform the machine learning algorithms to predict protests, in which they were successful in predicting future protests with an average prediction accuracy of over 75.

Compton et al.[3] perform the content analysis on tweets to find the important ones containing timestamps and location mentions of future protests in order to detect possible protests. Radinsky and Horvitz[8] use 22-year database and study the sequences of different events to predict a future event.

Steinert-Threlkeld et al.[10] also study the Arab Spring case using about 14 million tweets collected from 16 countries and show that there is a strong statistical relationship between protest activities in a particular day with the level of coordination in its previous day. Korkmaz et al. [5]design a prediction system (incorporated in EMBERS), and using data collected from Twitter and blogs in six Latin American countries, they show that the heterogeneous data sources can collectively increase the accuracy in prediction of future protests.

Mining a data stream in search of new patterns in the content of a document is essentially what amounts to a discovery challenge when it comes to the process of detecting an event. The purpose of this research is to make a prediction of public disturbance through the use of natural language processing (NLP) by utilizing the attributes extracted from the gathered tweets calling for protests. In particular, we look at the situation of the economic crisis that happened in Sri Lanka. There are two approaches to detect the sentiment of a tweet: (a) By locating indications of sentiment. (a) By analyzing the tweets to determine which side of an argument they take. In addition, the length of the tweets, the frequency of the tweets, the number of retweets, and the amount of hash tags are all factors that will assist determine the extent of the impact on society. The primary contribution of our work is twofold: first, the findings of our study provide evidence that Twitter analytics can successfully assist in the prediction of major protests; second, we demonstrate that the incorporation of regional characteristics that are driven by the causes and factors that motivate protests can assist in the production of more accurate predictions.

3 DATA COLLECTION

For the purpose of analysis, we have gathered data from Twitter. On "9th of July," in "Sri Lanka," the public of the nation marched on the president's house in the wrath of the economic crisis the country was suffering from at the time. This is the incident that we are trying to replicate. The time line that the data was collected for is broken down into three sections: the data collected before the event, which spans the period from 2 May 2022 to 2 July 2022; the data collected during the event, which spans the period from 3 July 2022 to 16 July 2022; and, finally, the data collected after the event, which spans the period from 17 July 2022 to 16 July 2022. The quantity of tweets that were collected varied from 40,000 tweets for pre-event analysis to 10,000 tweets during the event itself to 20,000 tweets for study after the event(Fig.1). We have combed through the Twitter feed looking for the most popular hashtags in that area during that time period. The

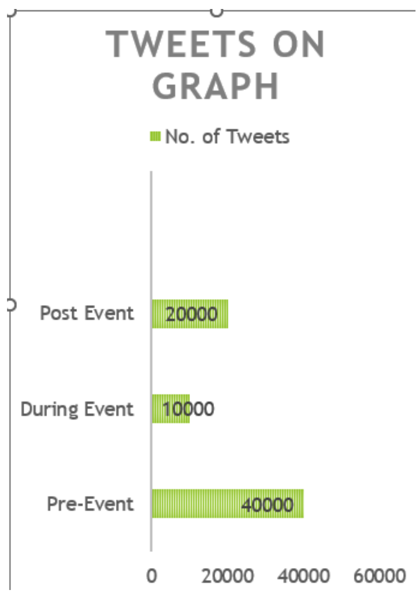

Fig. 1. Total Tweets

following is a list of the most common hashtags that we use: gotagohome OR gotagoranil OR srilankaeconomiccrisis OR gohomegota OR gohomemahinda OR gohomerajapaksas OR gogotago OR srilankacrisis OR lka OR fuelcrisislk OR srilanka demonstrations.

4 METHODOLOGY

In our project, we made use of the Web-Scraping methodology, which is a method for automatically extracting information by parsing hypertext tags in order to obtain plain text information that is embedded in the hypertext. With the help of the sncrape module, the dataset that comprised the links and tweet replies has been filtered out in order to obtain considerable data. After acquiring the data that we needed, we eliminated the stopwords and used the nltk package to iteratively process the regular expression tokenizer in order to refine the dataset even further. The CSV file that contains the original tweets has been updated to include the acquired tweets. After obtaining the most recent data, it was categorized so that a sentiment score could be calculated. The score for the negative category is zero, and the score for the positive category is one. Utilizing the Pandas dataframe allows for the sentiment function to be called, and an existing csv file is updated to include the newly added text. The sentiment analyzer is explained in Fig.2. The sentiment text is given a score, and that score, together with the date, is combined together using the aggregation function on the score. Invoking the plotting function allows the graphs to be constructed between the dates and the sentiment ratings that correspond to those dates when all of the data has been updated.

5 EVALUATION

When we look at the Pre-event dataset Fig.3 and Fig.4, we can see that the percentage of the sentiment of tweets was mostly negative and only slightly positive on the graph. This graph has deflections and is in a state of constant change as we get closer to the event; more specifically, the number of negative tweets is growing at a rate that is proportional

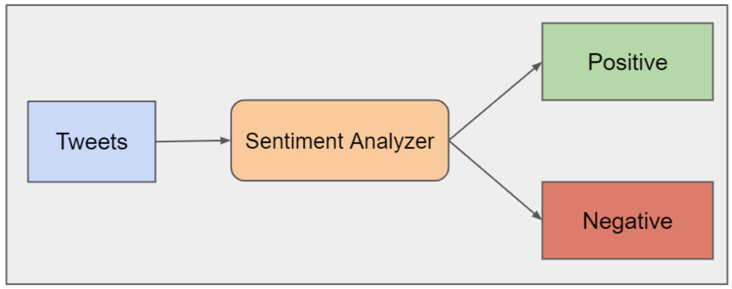

Fig. 2. Sentiment Analyzer

to the rate at which we are getting closer to the event. During the course of the event Fig.5 and Fig.6, the dataset reveals a relatively high percentage of negative tweets, while the percentage of positive tweets reaches saturation. This trend is reversing, albeit not to a significant degree Fig.7 and Fig.8, in the dataset collected after the event, as there has been an increase in the amount of favorable tweets while people have been cruising away from the event.

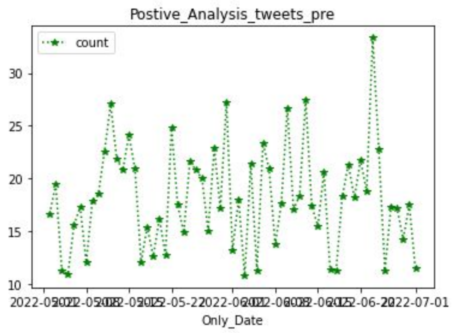

Fig. 3. Postive Tweets Pre-Event

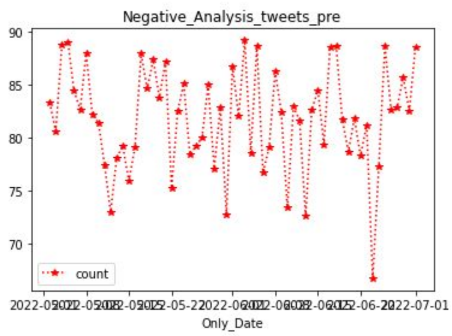

Fig. 4. Negative Tweets Pre-Event

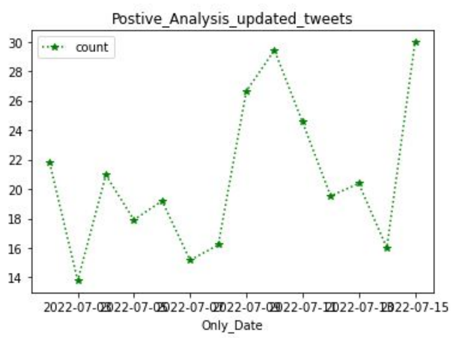

Fig. 5. Positive Tweets During-Event

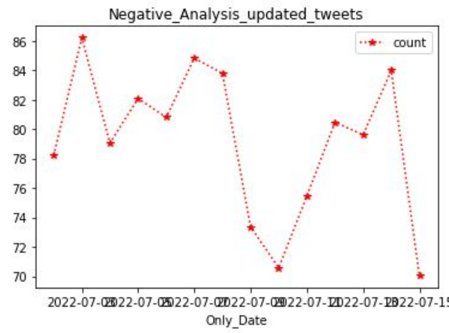

Fig. 6. Negative Tweets During-Event

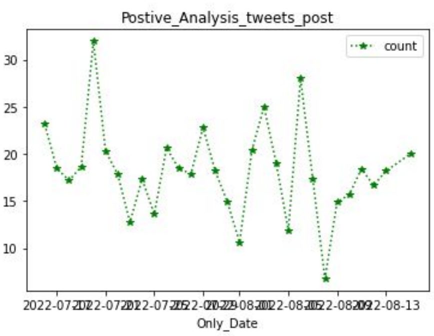

Fig. 7. Positive Tweets Post-Event

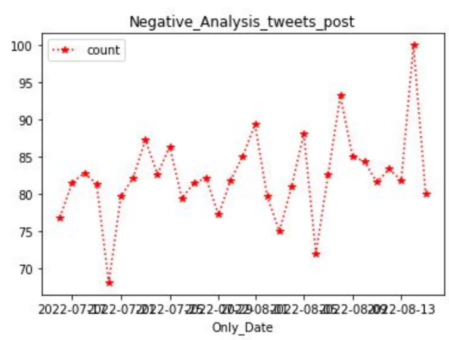

Fig. 8. Negative Tweets Post-Event

6 CONCLUSION

There is just a small amount of data that can be obtained from the twitter, and the authentication process for the user key takes an extremely long time. It is necessary to clean the data in accordance with the particular use case. The tweets take on a different tone once the stop words are eliminated. After the cleaning process, the relevant data set that is left over is significantly reduced in size compared to the original corpus. When tweets are written in a variety of languages, things get really difficult and complicated. It required a great deal of caution in order to avoid finding tweets that were irrelevant or repeated and contained the same hashtags. The study of event detection is an academic discipline in and of itself. We might include user behavior and do tweet analysis that is narrowed down to specific hours in order to determine whether or not there is a shift in the trend between morning and night. Ability to construct a twitter conversation analyzer for the purpose of gaining a more in-depth sentiment. Recognition of an entity can have a significant influence on determining how it can mobilize the sentiment of the general population. The sentiment of the tweets can be inferred with the use of part of voice recognition. The gathering of data can be done using a variety of methods, each of which will produce results that are more exact and precise.

In particular, we want to access the situation of the economic crisis that happened in Sri Lanka as well as the unrest that developed as a result of the crisis. In addition, we want to validate our model by forecasting that the riots that occurred on January 6 that occured during the transition of power in the United States presidency in 2021 and the war between Russia and Ukraine in 2022. We then make an attempt to explain that both instances of protesting likely brought about the same sentiment among the general population. Considering more events from the world will give a more accurate understanding of the sentiment of the public.

REFERENCES

[1] Mohsen Bahrami, Yasin Findik, Burcin Bozkaya, and Selim Balcisoy. 2018. *Twitter Reveals: Using Twitter Analytics to Predict Public Protests.* Technical Report.

[2] Kartikeya Bajpai and Anuj Jaiswal. 2011. *A Framework for Analyzing Collective Action Events on Twitter.* Technical Report. Pennsylvania, USA.

[3] Ryan Compton, Craig Lee, Tsai-Ching Lu, Lalindra De Silva, and Macy Michael. [n.d.]. Detecting future social unrest in unprocessed Twitter data:"Emerging Phenomena and Big Data". *Commun. ACM* ([n. d.]). https://doi.org/10.1109/ISI.2013.6578786

[4] Summer Harlow and Thomas J. Johnson. [n.d.]. Overthrowing the Protest Paradigm? How The New York Times, Global Voices and Twitter Covered the Egyptian Revolution. ([n. d.]), 16. https://ijoc.org/index.php/ijoc/article/view/1239/611

[5] Gizem Korkmaz, Jose Cadena, Chris J. Kuhlman, Achla Marathe, Anil Vullikanti, and Naren Ramakrishnan. [n.d.]. Combining Heterogeneous Data Sources for Civil Unrest Forecasting. ([n. d.]), 258–265. https://doi.org/10.1145/2808797.2808847

[6] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. [n.d.]. Twitter under crisis: can we trust what we RT? ([n. d.]), 71–79. https://doi.org/10.1145/1964858.1964869

[7] Panagiotis Panagiotopoulos, Alinaghi Ziaee Bigdeli, and Steven Sams. [n.d.]. "5 Days in August" – How London Local Authorities Used Twitter during the 2011 Riots. *Commun. ACM* ([n. d.]), 102–113. https://doi.org/10.1007/978-3-642-33489-4_9

[8] Kira Radinsky and Eric Horvitz. [n.d.]. Mining the web to predict future events. ([n. d.]), 255–264. https://doi.org/10.1145/2433396.2433431

[9] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. [n.d.]. Earthquake shakes Twitter users: real-time event detection by social sensors. ([n. d.]), 851–860. https://doi.org/10.1145/1772690.1772777

[10] Zachary C Steinert-Threlkeld, Delia Mocanu, Alessandro Vespignani, and James Fowler. [n.d.]. Online social networks and offline protest. *Commun. ACM* 4, 19 ([n. d.]), 9. https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-015-0056-y