JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-16, No.-7, July (2021) pp 22-31 ISSN (Print) 0973-8975

SENTIMENT ANALYSIS OF CURRENT TRENDING TOPICS ON TWITTER USER BASE

Zeeshan Rasheed[1], Naeem Ahmed Ibupoto[2], Syeda Surriya Bano[3] Sheeraz Ahmed[4]

[1]Department of Computer Science, Mir Chakar Khan Rind University Sibi, Pakistan

[2]Department of Computer Science, Mir Chakar Khan Rind University Sibi, Pakistan

[3]Department of Computer Science, Karachi, Pakistan

[4]Department of Computer Science, Iqra National University, Peshawar, Pakistan

Email: zeeshanrasheed1992@yahoo.com, naeemloher@gmail.com surriya.bano138@gmail.com

Corresponding Author: **Zeeshan Rasheed**

https://doi.org/10.26782/jmcms.2021.07.00003

(Received: May 8, 2021; Accepted: July 3, 2021)

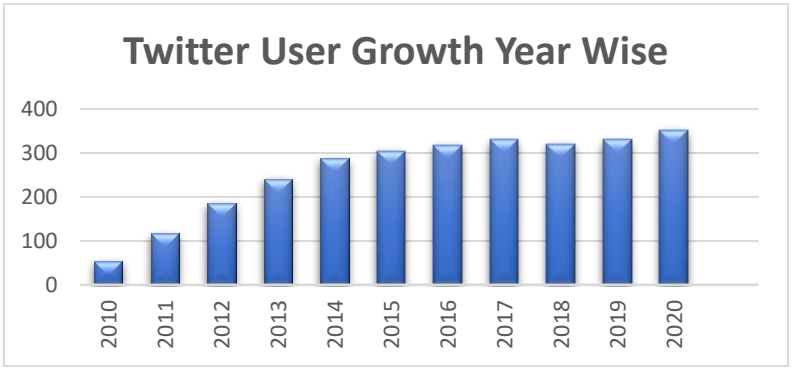
Abstract

Twitter has now become the most common social platform to express views on any topic. A micro-blogging social media offers a way for people around the world to show their sentiments about any political, social and cultural subject of the time. In this paper, the sentimental analysis approach has been used to analyze the positive and negative sentiments of Twitter users about some top trending #tags around the globe. The data has been collected between the duration of March to April 2021. The collected data were processed by using the Python program and then transformed our data set with the help of the SQL database. We have used graphs and tables to present the data, collected under three hashtags; which were the top trending topics in that particular era. The tweets were elaborated by positive, negative and neutral sentiments which were depicted in graphs. It is clear from the results and comparison that social media has a strong influence in the present era and can be highly helpful to use as a predictor of any political, social situation prevailing in any country or worldwide. It has also been helpful for business communities to analyze their products in the same manner to improve their business growth.

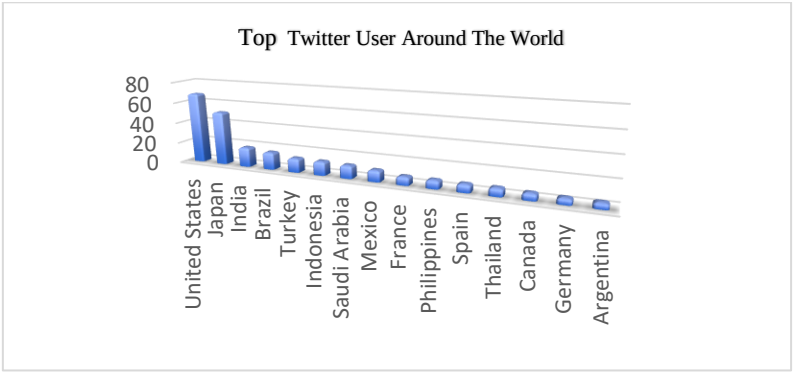
Keywords: social platform, social media, #tags, SQL, SA, API

I. Introduction

Twitter, a microblogging site, created in 2006 has now become a social platform to express views on any social and political issue. Within a span of seven years, it has become one of the most visited websites. More than 300 million active users post, comment and express themselves on Twitter. Tweets and replies to tweets can be sent by cell phones, text and work region clients, or by posting on the site with a maximum limit of 140 characters [I].



Graph: 1 Twitter User Growth Year Wise



Graph: 2 Top Twitter Users Around the World

Sentiment Analysis is an approach used, in analyzing the sentiments of Twitter users about political or social topics. These sentiments can be positive, negative, or sometimes neutral. Words and phrases used in tweets express the sentiments and emotions of people with particular mindsets. Machine Learning (ML) is a technique that uses the artificial intelligence of machines to analyze and calculate data for future predictions. In business and other affairs, this is also known as prophetic research. The purpose of this research is to analyze the views and sentiments of Twitter users on top trending social issues of the world in a specific time frame. We have tried to relate the outcomes of our study with real-life by keeping in mind the objectives and purpose of our research, related to the sentiments of Twitter users about the scenario world regarding these topics The tools and software used; are Python program to collect and extract tweets related to our topic. Twitter API and SQL Lite 3 were used for establishing datasets and filtering tweets as English and non-English. Excel and Chorus were used for analyzing and arranging the collected data.

We also faced some challenges during this process of data collection that hamper the work i.e. the extensive use of Roman and Urdu tweets and abbreviations like lol, hahahahah, oh and many other emojis.

After the analysis of our findings, it was clear that in future work, we would also consider the Roman and Urdu tweets for much accurate and broad study in this field.

II. Literature Review

The research work related to sentiment analysis is not less; as there is various research work that investigates Viewpoint Analysis (SA) [II]. Supposition investigation (SA) is generally a new term in this field. Twitter with its limit of 140 – a character for tweets allows its users to be precise in their emotions expressed. The best outcomes got in the characterization of emotions utilizing managed learning methods, for example, Naive Bayes and Support Vector Machines; however, the manual coding vital for the regulated methodology is extravagant. There has been some non-directed work approaches referenced in and semi-managed procedures that can be improved with further research [III], [V]. Specialists, with new qualities and evaluating strategies, frequently contrast their outcomes and starting execution. It is important to make formal correlations between these outcomes that get from various grouping qualities and methods to choose the best attributes and the most proficient characterization systems for the applications [VI].

Through this recent decade, the measure of work carried out in the sentiment analysis field has a generous increment. The work associated with Sentiment Analysis (SA) has generally taken after the accessible information, equally as far as the work is done and the concentration region like Opinion Mining, Sentiment Investigation, Opinion Characterization and Micro-Blogging. [VIII]

Moreover, it is essential to standardize the words stretched, for example, happyyyyyy, cooooooll and soooooooo replacing the letters which usually occur in the sequence with perhaps a couple of letters [IX].

III. Research Methodology

We conducted our study by gathering data from Twitter using Twitter API. It is not a new idea but we have to separate English tweets from other languages' tweets, but it was not an easy task. Manual work was also done for this purpose. The methodologies used for analysis were unique with research. We collected our data for research by keeping in mind the purpose of our study.

Data collection is not an easy process as it might appear at the initial stage. We used the keyword #hashtags that were further passed on Python using Twitter API and credentials. The data extracted from Twitter was of many languages but we focused on English tweets, gathered and transformed them into the dataset. To filter non-English tweets, we used SQL query to filter out all those tweets. After filtering tweets, we created a dataset to classify those gathered tweets into positive and negative. On the other hand, we ignored neutral tweets because they had no impact on analysis and results. Tweets were extracted by using the Python Program. It streams the data with the given hashtag. After applying for the program, Twitter started the collection of tweets against the passed # tags. SQLite is an associated title in the collection of procedures that uses an autonomous transaction SQL information machine, without a server, zero configurations. SQLite is the most implemented information in the world with additional applications that we are about to tell, along with many other high profile [IX]. It is the most implemented information, embedded level SQL without server method.

There is a commutation between speed and usage of memory SQLite generally runs the additional memory that is provided to it. However, it can perform quite well, even in low-memory environments. Betting on, however, SQLite is quicker than the direct file system. After passing the hashtag, data was streaming. When the tweets stop streaming the data was saved into database SQLite.

In the very next phase, we started converting our collected data in tabular form by using SQL Lite 3 database. The tweets were in raw form on which we had to elaborate the other non-required fields such as follower count, friends count, status count and other more elements from our dataset.

Up to this point, we were not sure about negative and positive tweets, the datasets that would be utilized for preparing and analyzing the tweets were of English and non-English.

As earlier discussed, we are focusing on English tweets only but the Twitter users from all over the world use different languages to express their sentiments and opinions. For this purpose, we need to filter other language tweets with SQL Query because in this study our key focus is on English language tweets.

We collected the tweets from March to April 2021 using Twitter API about five top trending and popular topics of Pakistan and worldwide. The process of tweets collection is done by passing the #tag to the transmission API. We can define the hashtag as a word that uniquely identifies any topic on Twitter. We gathered tweets against each #tag which was mentioned below.

A. Top Trending Topics around the Globe (March – April 2021)

We have collected three top trending topics around the globe table 3.1 shows the hashtags on which Twitter users around the world express their sentiments at the time of data collection. Then we plot a graph fig 3.1 below that shows the number of tweets collected for each topic under a specific heading; Topic Wise Tweet Collection.

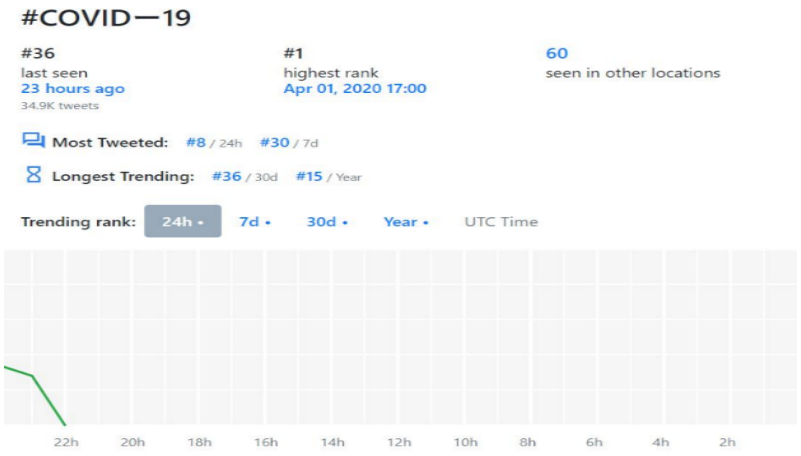


Figure: 1 #COVID-19

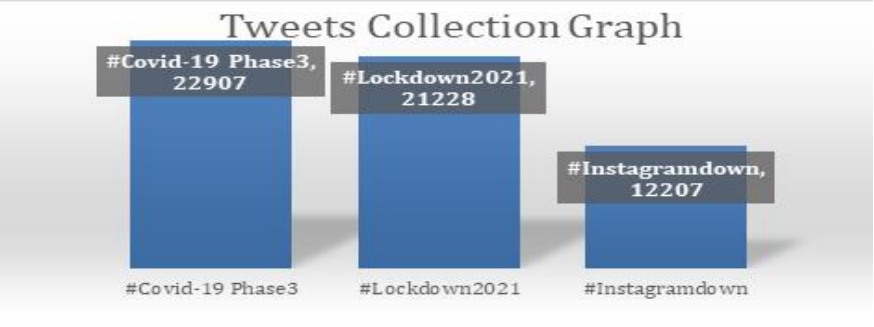
It is clear from figure 1 #COVID-19 (phase 3) was trending on the top on April 01, 2021

Figure: 2 #Lockdown2021

In figure 2 #lockdown our second topic was top trending on March 20, 2021. Many users were tweeting about it at the same time.

Table- 1: Top Trending Topics around the Globe

S.no	Top Trending Topics around the Globe (March – April 2021)	
	Trending Topic	***Ranking at the time of Collection***
1.	Covid-19 (Phase 3)	1st
2.	Lockdown	1st
3.	Instagram down	3rd



Graph: 3 Topic Wise Collection of Tweets

IV. Results and Discussions

The results show which were just predictions before; after analysis, those were outcomes to prove the sentiment analysis of Twitter users about top trending political issues. After the data collection and pre-processing, each tweet was classified as negative or positive bases on the polarity of words used. The polarity ranges from +5

to -5 where the higher value of positive was indicative of the very positive and the highest value of negative was indicative of very negative tweet. The value 0 indicated a neutral tweet, which was neither positive nor negative. The blue and red lines were used for the positive and negative tweets respectively. The terms positive or negative are relative terms and their definition varies from hashtag to hashtag.

All the labeling of tweets is done by polarity score using chorus sentiment tool i.e. we will first check the relevance of tweet to our definition and if a tweet falls into our definition then in the next phase we will use chorus for its scoring and labeling the tweets by giving score numbers. We also validated this method for performing manual analysis of randomly picked tweets from each hashtag.

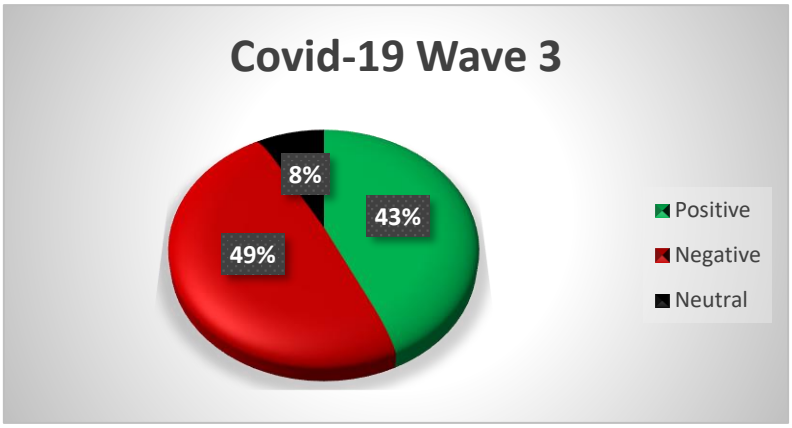
A. Covid-19 Phase 3

A new type of Corona Virus was first reported in Wuhan, Hubei province, China in Dec 2019. This disease has now spread around the whole world and with time, it is now converted into a pandemic. A pandemic is an epidemic of an infectious disease that has spread across a large region, for instance, multiple continents or worldwide, affecting a substantial number of people.

Symptoms of this virus are quite variable, but most of the time includes cough, fatigue, fever, difficulties in breathing, and senses of taste and smell are also affected. These symptoms may start between one to fourteen days after having any exposure to this deadly virus.

This Coronavirus spreads widely when a person infected with it, has any close contact with others. Till now, we have faced the three periods of this deadly disease. The first period, between 15th March and 30th June, corresponding to the entire first wave, and the second, between 1st July and 15th October. The third wave of this pandemic is still going on with its disasters.

The third wave of Covid-19 is deadlier than before and it has created widespread panic and alarm in the world after its detection. It is time to be more careful and protective now. We have chosen it as one of the top trending topics of the time. People have expressed their sentiments as positive, negative, or neutral through tweets. We have collected the data and presented it in the graph. The green color part shows the positive sentiments of the people about this new wave of Corona Virus as they are considering it seriously. The red color pasty shows the negative sentiment of the people as in their opinion this situation is not much critical or deadlier than before. While the black color part shows the number of people who have not bothered to express their views about this situation or they is neutral in this #tag. The trend shows that people are not taking this third wave seriously and are not cautious about it. The graph is shown below which clearly depicts the sentiments of Twitter users against this #tag.



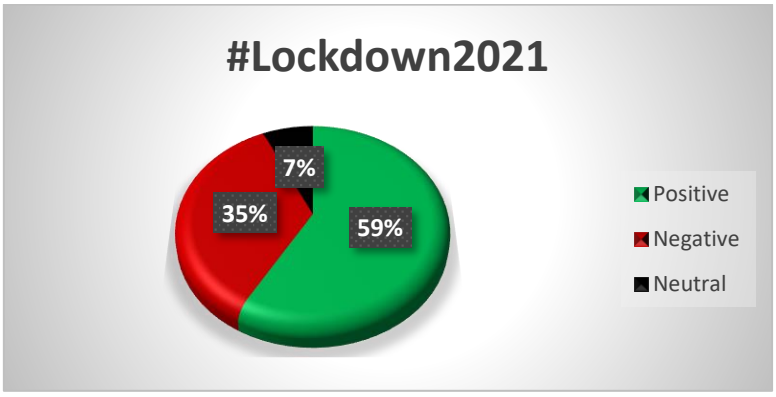
Graph: 4 Covid-19 Wave 3

B. #Lockdown2021

This hashtag has a close relevance with the previous trend Covid-19 third Wave. The whole world is now facing the third wave of Covid-19. This problem first arose at the end of 2019 and the start of 2020. This deadly disease soon spread like a wildfire in the whole world and millions of people became victims of it. As this virus was spreading with the close contact of the people so the governments all over the globe thought to adopt the lockdown policy as a necessary solution. The whole world was in a fix and the social and economical life of the people was stuck.

As the flow observed in the cases of Covid-19 around the world with a new wave of it, many countries have decided to re-impose the situation of lockdown to restraint this spread. This time the situation is more critical than the other two waves of this pandemic. Though vaccination has been started in many countries, yet the cases of victims of this disease are still increasing. Many European countries like France, Poland, Hungary, Belgium, Italy, the Philippines and Asian countries like Pakistan, India, Bangladesh with the United States of America and England have also gone to lockdown partially and a complete lockdown in some parts of these countries with the severity of the situation.

This situation has again influenced the lives of the people and they are taking it seriously with their routine life. According to our study, 59% of the people are in favor of lockdown to avoid any devastating situation. On the other hand, 35% of people are not in the favor of lockdown as it affects their normal daily routine. Only 7% are neutral respondents which shows the importance of this issue related to the lives of every single and coo on the person.



Graph: 5 #Lockdown2021

C. #instagramdown

Instagram is a social networking platform, created by Kevin Systrom and Mike Krieger on October 6. It is mostly used to share photos and videos. In April 2012, Facebook acquired the service for US$1 billion approximately in cash and stock. This app has the features of editing and sharing the media with filters and it can be organized by #tags and geographical tagging. A person can share the posts publically or which preapproved followers. Users have the advantage to browse the contents of other users by specific tags and locations and check also new trending content. Users have the option to like and follow the photos and contents of other users and they can also add their content to a personal feed. [XI]

Instagram was suddenly down on March 20, 2021, and thousands of people were unable to post and see their feeds globally due to this problem. It again faced the same problem after some days. Though this problem was resolved in less than an hour people started sharing their sentiments and views through words and memes on Twitter.

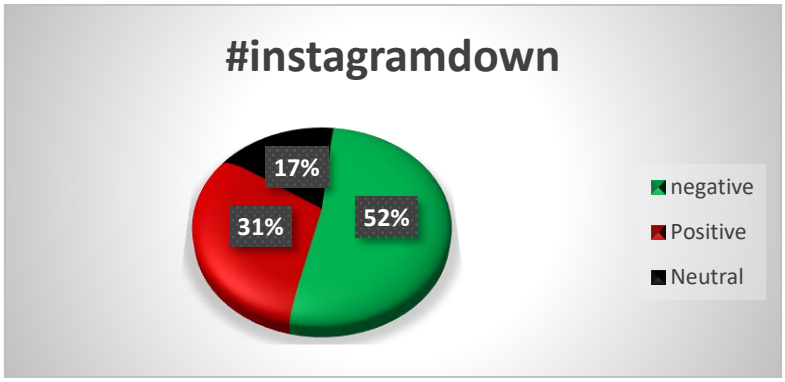
It is too early to figure out what the reason could be behind the crash. Even Instagram (or Facebook) has not acknowledged such an issue is impacting people. But this crash seems to be more intense than expected. According to people, the Instagram app keeps crashing even after it has been uninstalled and reinstalled on the phone. Rebooting the phone is also not helpful in the situation. So, it seems there is no way to circumvent the crash problem until Instagram takes hold of the matter. The reasons for this problem are still not discussed by the Instagram authority. People are just trying to guess some issues but they can't be taken for sure.

Somehow, it was a top trend on Twitter and we considered it as our topic and collected tweets accordingly.

The graph of sentiment analysis shown the Twitter user behavior towards this subject. The green part shows that 52% of opinions were with the positive tendency. People expressed their opinions that due to some unknown technical problem, Instagram was down and we have to cooperate and they also suggested shifting to some other social communication app for a while and in this way they can keep in touch with their friends and family and social groups around the world.

The red part is 31% which shows the negative sentiments of the people about this problem. People expressed their sentiments negatively as insta goes down and they can't able to share their views and make contact with their circle around the globe as it was a worldwide issue and ranked no 1 in top trends of Twitter. A chart depicts this situation very clearly.

Only 17% of people were neutral regarding this issue as Instagram down had not affected their social life or they did not consider it seriously.



Graph: 6 #instagramdown

V. Conclusion and Future Work

Sentiment analysis has now become an important source in decision-making. People depend upon it for future predictions and efficient working, yet we cannot say it is more than enough to rely upon its judgments because opinions and perceptions of people vary with the circumstances. Overall, we can predict or make up our minds for the future according to people's sentiments either positive or negative.

In this paper, we have used the Sentiment Analysis (SA) approach to predict the people's views about three top trending #tags around the world in a specific time period. Our findings show that social media like Twitter helps to establish a sound perception of the social, political and cultural issues by analyzing the thoughts and feelings of people concerning their comments. Our research shows that only 5.43% of people commented in the English language; rather people prefer to comment in their native languages or Roman Urdu and abbreviations. In future work, Roman Urdu and other languages can be also included for broader research and analysis. These types of studies and research open new vistas for researchers to further explore this field of sentiment analysis of all Twitter users irrespective of any language usage. As the languages in tweets and retweets are more, the studies can be carried out in a wider sense. This study is also a base for further work in sentiment analysis of social media users about any other famous topic or notorious. This type of study is also helpful for business communities on which they can easily analyze their products and keep themselves updated.

In addition, this research can make some sound grounds for further predictions by using social media platforms like Twitter.

Conflicts of Interest:

The authors declare that they have no conflicts of interest to report regarding the present study.

References

I. A. Giachanou and F. Crestani, 'Like It or Not: A Survey of Twitter Sentiment Analysis Methods,' ACM Comput. Surv., vol. 49, no. 2, pp. 1-41, 2016

II. Diakopoulos, N. & Shamma, D., 2010. Characterizing Debate Performance via Aggregated Twitter Sentiment. Proceedings of the 28th int. conference on Human factors in computing systems. Go, A., Bhayani, R. & Huang, L., 2009. Twitter Sentiment Classification using Distant Supervision."

III. https://en.m.wikipedia.org/wiki/Instagram

IV. Koyel Chakraborty, Sudeshna Sani, Rajib Bag,: 'A STUDY ON SENTIMENT POLARITY IDENTIFICATION OF INDIAN MULTILINGUAL TWEETS THROUGH DIFFERENT NEURAL NETWORK MODELS'. *J. Mech. Cont.& Math. Sci., Vol.-15, No.-1, January (2020) pp 108-117*. DOI : 10.26782/jmcms.2020.01.00008.

V. Larsen, Peder Olesen, and Markus von Ins. 'The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index.', Scientometrics 84.3 (2010): 575–603. PMC. Web. 25 Sept. 2015."

VI. Liu, Bing, and Lei Zhang. 'A survey of opinion mining and Sentiment Analysis (SA). Mining text data. Springer, Boston, MA, 2012. 415-463."

VII. Omar bin Md Din, Abdul Ghani Bin Md Din, Rusdee Taher, Abduloh Usof, Prasert Panprae, Yousef A. Baker El-Ebiary. : 'WEB CONTEXT AND THE MULTIPLE SEMANTIC LINGUISTIC ORIGINS AND ITS IMPACTS ON THE PROPHET'S TEXT. *J. Mech. Cont.& Math. Sci., Vol.-15, No.-7, July (2020) pp 392-404*. DOI : 10.26782/jmcms.2020.07.00033.

VIII. Polk, Alexander, and Patrick Paroubek. 'Twitter as a corpus for Sentiment Analysis (SA) and opinion mining.' LREc. Vol. 10. No. 2010. 2010."

IX. Selmer, Øyvind, et al. 'NTNU: Domain semi-independent short message sentiment classification.' Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Vol. 2. 2013."

X. Turney, Peter D. 'Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews.' Proceedings of the 40th annual meeting in association for computational linguistics. Association for Computational Linguistics, 2002."

XI. Twitter - Wikipedia.