

Twitter Sentiment analysis during COVID-19 Outbreak in Nepal

Bishwo Prakash Pokharel

M.Phil in ICT 3rd semester

Nepal Open University, Nepal

Abstract

The growth of social data on the web is increasing rapidly nowadays. This leads to researchers to access the data and information for the research purpose. During the global COVID-19 outbreak, many individuals as well as organizations and government agencies are posting their viewpoints regarding the coronavirus. The study focuses on the sentiment analysis of tweets of the Twitter social media using Python programming language with Tweepy and TextBlob library. The tweets have been collected, pre-processed and then used for text mining and sentiment analysis using Google Colab. The graphical representation has been provided on the data after sentiment analysis based on two specified hashtags keywords : #COVID-19 and #coronavirus. The data are collected from the users who shared their location as 'Nepal' between 21st May 2020 and 31st May 2020. The result of the study concluded that while majority of the people of Nepal are taking a positive and hopeful approach, there are instances of fear, sadness and disgust exhibited too.

Keywords

COVID-19, coronavirus, TextBlob, Tweepy, Sentiment Analysis, Twitter

1. Introduction

Nowadays the social media platforms such as Twitter, Facebook and YouTube are a great source of information known as social data. Peoples are sharing their news and viewpoints in the social media and discuss about the posts between their friends, relatives and the people on their virtual connection with their accounts and express their opinion about these events with positive, negative or neutral reaction on that post. The origin of COVID19 is said to be in the starting of December 2019, when several patients from Wuhan, Hubei Province reported severe respiratory infections. These patients had background of working in the wholesale fish and seafood market, also known as wet markets[1].

In Nepal, the first case was reported on 23 January 2020, a 32-year old Nepali man returning from Wuhan. The patient recovered and contacts were also asymptomatic. Immediate action were taken to strengthen the health desks at Tribhuvan International Airport and gradually at other airports as well. The ground crossing Points of Entry (PoE) at the Nepal-China border and the Nepal-India border were strengthened with health desks. Traffic limitations are in place on both sides of the borders. The Nepal-China official border crossing points have remained closed since 21 Jan 2020. The Government of Nepal announced suspension of all international flights followed by a country-wide full lockdown since 23 March 2020. The GoN repatriated 175 Nepalese from six cities across Hubei Province of China on 15 February. They were kept in quarantine in Kharipati, Bhaktapur and no cases were confirmed from within that group[2]. As of 06 June 2020, 92477 RT-PCR tests have been performed and 3235 cases have been found to be positive. Among them, 13 death and 365 cases have been discharged from hospitals and average hospital stay of COVID-19 discharged case is 16 days[3].

World Health Organization has suggested that isolation and self-quarantine is one of the major ways to stop this pandemic to spread with such an alarming rate. Nepal has witnessed benefits of the one of the countrywide lockdown at the start of this pandemic. As of 06 June 2020, 168031 people are in Quarantine and 2857 confirmed cases are in isolation[3].

The social media is revolving with the people to get them involved by making current trending insights concepts that is trending within a second .In the recent years, the people are exposing their social related issues through several social media by comments, reviews, posts, hashtags, emoji's etc. which was followed by many people and those tweets become popular soon[4]. The contents which are posted and discussed on the social media is noticed and can be used for the analysis purpose. The Sentiment Analysis (SA) is used to analyze the information and the reaction to every aspect can be clearly analyzed with three aspects that is positive, negative or neutral.

This research has been done to identify the sentiment of the Nepalese citizens regarding the corona virus outbreak. In this paper, tweets have been taken from Twitter using Twitter API with the search keyword specified as #coronavirus and #COVID-19 and sentiment analysis has been performed on the basis of that tweets who shared their location as Nepal from the date 21st May to 31st May 2020. Python programming language has been chosen because it provides many easy to use libraries to collect, process and visualize the data. All the data are gathered by using tweepy library and sentiment analysis is performed by using TextBlob library of Python in Google Colab.

2. Problem Domain

Sentiment analysis in the domain of micro-blogging is a relatively new research topic and decent amount of related prior work has been done on sentiment analysis of user reviews, documents, web blogs, articles and general phrase level sentiment analysis. These differ from twitter mainly because of the limit of 280 characters per tweet which forces the user to express opinion compressed in very short text. This research intends to collect, process and analyze the sentiments of people on the tweets posted on twitter during the certain time frame. So, the research will focus on the following questions:

- Collect the tweets through Twitter API using tweepy library of Python with hashtags #COVID-19 and #coronavirus.
- Preprocess the tweets by data cleaning (removing white spaces, links, punctuations, stop words, tokenization)
- Calculate the sentiment using TextBlob library and analyze the result.

3. Method for Analyzing the Papers

The sentiment analysis of Twitter data is an emerging field of Natural Language Processing in Artificial Intelligence. In this paper, the process of data collection is done using Tweepy python library with the Twitter API provided by the Twitter Developer. From this API various fields like text, source, retweets, language, user, location etc. are scrapped. After collecting data TextBlob library is used for finding the polarity of the text (positive text, negative text or neutral text). Sentiment analysis has a certain procedure which start by collecting data, identify the data and required features are extracted. After that sentiment classification is performed and finally analysis is done on the basis of sentiment polarity and subjectivity as in following figure:

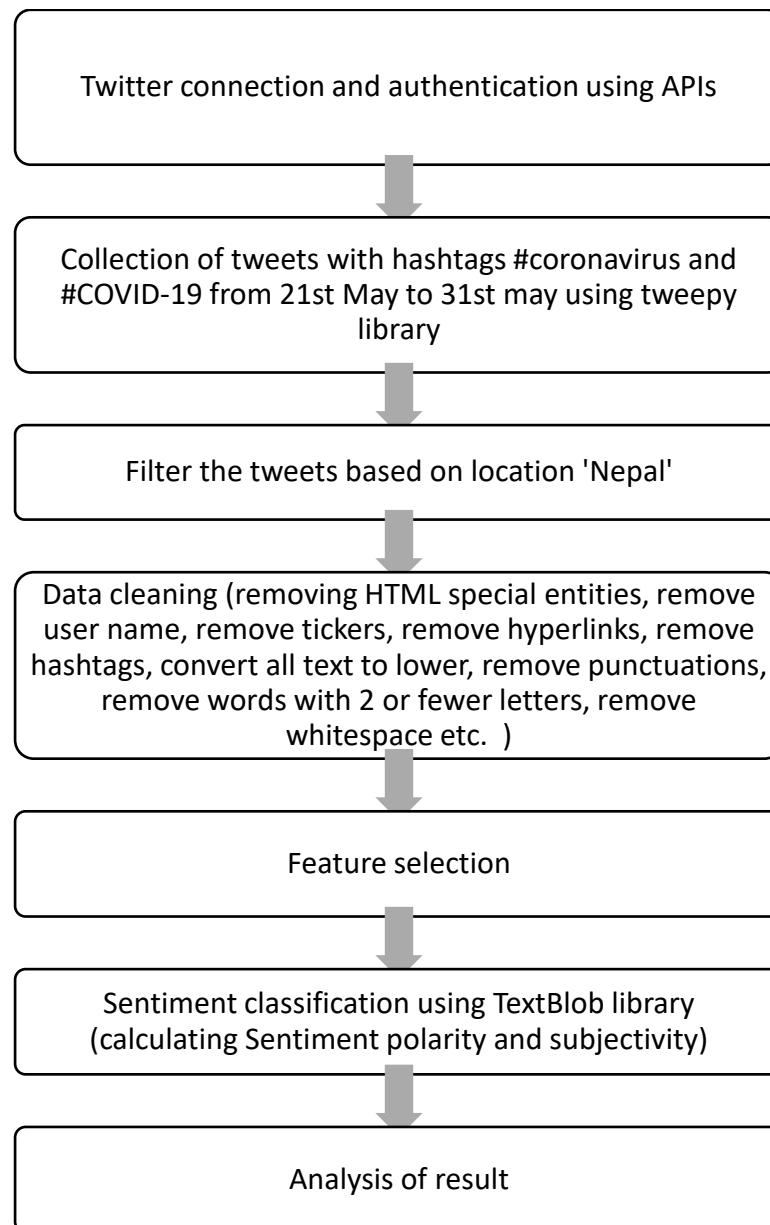


Figure: Methods of analyzing the research paper

4. Literature Review

Sentiment analysis has been widely researched in the domain of online review sites with the aim of generating summarized opinions of users about different aspects of products. Identifying such sentiments from online social networking sites can help emergency responders understand the dynamics of the network, e.g., the main users' concerns, panics, and the emotional impacts of interactions among members[5]. Advent of social media has created an unprecedented environment for people to share their thoughts with the world. These online platforms like Facebook, twitter are usually the

first resort people turn to in times of crisis to voice their opinions and relay other crucial information. But when it comes to detecting sentiments out of this gigantic pool of opinions, it becomes an arduous task and doing it manually is practically impossible. Hence different methods have been devised to perform automatic polarity classification of textual data[6].

The most prominent of these methods rely on two approaches: lexicon-based and learning-based. In the first, the analysis of a document's expressed sentiment is achieved through its breakdown to words whose sentiment polarity is pre-defined in a lexicon. In learning-based, supervised classification algorithms are fed with pre-annotated (in terms of their sentiment polarity) documents, and are trained in order to autonomously classify future inputs[7].

Opinion and sentimental mining are important research areas because due to the huge number of daily posts on social networks, extracting people's opinions is a challenging task. Sentiment analysis has many applications for different domains for example in businesses to get feedbacks for products by which companies can learn users' feedback and reviews on social media to analyze the customer review[8]. There has been a lot of effort been put in this field where programmers have applied soft programming approaches, That is usually fuzzy logic and neural networks for sentiment analysis. There are algorithms construct fuzzy domain sentiment ontology tree based on the reviews that includes the extraction of sentimental words or sentences, distinct features of the products and relation amongst features thus precisely predicting the polarity of the reviews in the networking site. By designing membership functions for the process they formulated and standardized the elite process of evaluating the strength of reviewer's opinions in the presence of an adverbial modifier on the social networks[9].

The data are taken from Twitter using Tweepy API for analyzing the respective data by the use of sentiment analysis. The data has to be cleaned before any preprocessing technique can be applied to the same. The topic can be mentioned in different contexts like a tweet or a retweet or a hashtag for example. In some other related studies tweets have been carried out to evaluate the preferences of users over a certain class of products. Political Sentiment analysis can be carried out before elections to know which party has the highest chances of getting the majority.

5. Analysis of Research Works

Many researchers are working on Sentiment Analysis on Twitter in literature, some key contributions which are providing support for finding user behaviors and situations in the different cases while happening around the world.

A. Bastola, R. Sah, A. Rodriguez-Morales et al.[10] has discussed and explained the detail case about the first 2019 novel coronavirus case in Nepal. They have been deeply studied with the symptoms and all the consequences with the initial radiograph of the patient.

R. Muthausami, A. Bharathi, K. Saritha [11] has done research on the topic COVID-19 Outbreak: Tweet based Analysis and Visualization towards the Influence of Coronavirus in the World. They have utilized machine learning methodology as part of a request to give a careful examination of the tweets which have been extracted from Twitter. Tweets have been classified into positive, negative and neutral classes with the help of the different machine learning classifier as Naïve Bayes, SVM, Decision Tree, MaxEntropy, LogitBoost and Random Forests. In the experiment with three classes, the LogitBoost ensemble classifier gets the most noteworthy accuracy.

C. Kaur, A. Sharma[8] has done Twitter Sentiment Analysis on Coronavirus using Textblob library. In this research paper, the opinion of different peoples of different countries on coronavirus the affect various countries has been discussed. The main focus of this paper is on Twitter, Twitter API and have implemented the python programming language and write code in Jupyternotebook to implement the sentimental analysis as positive, negative and neutral. The results are shown by using Matplotlib library.

M. Lwin, J. Lu, A. Sheldenkar et al.[12] has done research on the topic Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. They have collected over 20 million social media twitter posts made during the early phases of the COVID-19 outbreak from January 28 to April 9, 2020 using “wuhan”, “corona”, “nCov” and “covid” as search keyword and analyzed the emotions of tweets using the algorithm CrystalFeel. Pearson r correlations were conducted

between emotions across time statistically. Word clouds were generated for each of the four emotions based on the top frequent unigrams and bigrams.

J. Samuel, G. Ali, M. Rahaman et al.[13] in their paper has identify public sentiment associated with the pandemic using Coronavirus specific Tweets and R statistical software, along with its sentiment analysis packages. They demonstrate insights into the progress of fear-sentiment over time as COVID-19 approaches peak levels in the United States, using descriptive textual analytics supported by necessary textual data visualizations.

N. Rajput, B. Grover, V. Rathi[14] presents a statistical analysis of the twitter messages related to this disease posted since January 2020. Two types of empirical studies have been performed. Unigram, bigram and trigram frequencies have been modeled by power law distribution. The results have been validated by Sum of Square Error (SSE), R^2 and Root Mean Square Error (RMSE). High values of R^2 and low values of SSE and RMSE lay the grounds for the goodness of fit of this model. Sentiment analysis has been conducted to understand the general attitudes of the twitter users at this time.

S. Zivanovic, J. Martinez, J. Verplanke [14] has explores the use of social media in Quality-of-Life (QoL) research by capturing and mapping people's perceptions about their life based on geo-located Twitter data. The methodology is based on a mixed-method approach, combining manual coding of the messages, automated classification and spatial analysis.

A. Dubey [1] has done the country wise sentiment analysis of the tweets which has taken into account the tweets from twelve countries. These tweets have been gathered from 11th March 2020 to 31st March 2020, and are related to COVID19 in some or the other way. This analysis has been done to analyze how the citizens of different countries are dealing with the situation. The tweets have been collected, pre-processed, and then used for text mining and sentiment analysis.

H. Drias, T. Houari, Y. Drias [15] has collected a dataset of more than 600,000 tweets with hashtags like #COVID and #coronavirus posted between February 27, 2020 and March 25, 2020 was built. An exploratory treatment of the number of tweets posted by country, by language and other parameters revealed an overview of the apprehension

of the pandemic around the world. A sentiment analysis was elaborated on the basis of the tweets posted in English because these constitute the great majority (USA, GB, India...). On the other hand, the FP-Growth algorithm was adapted to the tweets in order to discover the most frequent patterns and its derived association rules, in order to highlight the tweeters insights relatively to COVID-19.

O. Gencoglu, M. Gruber [16] has propose a casual inference approach to discover and quantify casual relationships between pandemic characteristics (e.g. number of infections and deaths) and Twitter activity as well as public sentiment.

R. Medford, S. Saleh, A. Sumarsono et al. [17] extracted tweets matching hashtags related to COVID-19 and measured frequency of keywords related to infection prevention practices, vaccination, and racial prejudice. They have performed a sentiment analysis to identify emotional valence and predominant emotions by using topic modeling to identify and explore discussion topics over time.

A. Abd-alrazaq, D. Alhuwail, M. Househ et al.[18] has studied a set of tools (Twitter's search application programming interface (API), Tweepy Python library, and PostgreSQL database) and using a set of predefined search terms ("corona," "2019-nCov," and "COVID-19"). They have extracted the text and metadata (number of likes and retweets, and user profile information including the number of followers) of public English language tweets from February 2, 2020, to March 15, 2020 and analyzed the collected tweets using word frequencies of single (unigrams) and double words (bigrams). They leveraged latent Dirichlet allocation for topic modeling to identify topics discussed in the tweets. They have also performed sentiment analysis and extracted the mean number of retweets, likes, and followers for each topic and calculated the interaction rate per topic.

Mohammed Alhajji, A. Khalifah, M. Aljubran et al.[19] used Naïve Bayes machine learning model to run Arabic sentiment analysis of Twitter posts through the Natural Language Toolkit (NLTK) library in Python. Tweets containing hashtags pertaining to seven public health measures imposed by the government were collected and analyzed.

R. Kaila, A. Prasad [20] has done study focuses on the information flow on twitter during the corona virus outbreak. Tweets related to #coronavirus are studied using

sentiment analysis and topic modelling using Latent Dirichlet Allocation post preprocessing. The study concluded that the information flow was accurate and reliable related to corona virus outbreak with minimum misinformation. LDA analysis had identified the most relevant and accurate topics related to corona virus outbreak and sentiment analysis confirmed the prevalence of negative sentiments like fear along with positive sentiments like trust.

K. Suppala, N. Rao [21] has tried to perform sentiment analysis on twitter data by using a Naïve Bayesian algorithm. By using this model, they have measured the customer opinions and perceptions.

K. Manguri, P. Amin [22] has pulled out twitter data from Twitter social media, through python programming language using Tweepy library, then by using TextBlob library in python the sentiment analysis operation has been done. After the measuring sentiment analysis, the graphical representation has been provided on the data. The data they have collected on twitter are based on two specified hashtag keywords, which are “COVID-19” and “coronavirus”. The date of searching data is seven days from 09-04-2020 to 15-04-2020. In the end a visualized presentation regarding the result and further explanation are provided.

6. Implementation of the research project

Tweepy python library has been utilized for data extraction from Twitter API. Tweepy allows appropriate data retrieval by searching via keywords, hashtags, timelines, trends, or geo-location[23]. However, this research is targeted to Nepal only for data collection. In spite of having many restrictions from Twitter API, successive attempts to access as many posts as possible have been applied.

All the tasks have been done in Google Colab. Google Colab is a free cloud service which supports free GPU ! We can improve our python programming language coding

skills and develop deep learning applications using popular libraries such as Keras, TensorFlow, PyTorch, and OpenCV[24].

Following necessary libraries have been imported:

```
[ ] import csv
import os
import pandas as pd
import tweepy
import re
import string
from textblob import TextBlob
import preprocessor as p
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
```

After importing the library, Twitter API has been created with the following attributes and passed the credentials to tweepy:

```
[ ] #Twitter credentials for the app
consumer_key = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX'
consumer_secret = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX'
access_key= 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX'
access_secret = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX'
```

```
[ ] #pass twitter credentials to tweepy
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_key, access_secret)
api = tweepy.API(auth)
```

After that, different columns has been set and data are collected on the basis of start date and end date:

```
#set two date variables for date range
start_date = '2020-05-21'
end date = '2020-05-31'
```

While collecting the data non English keywords, stop words, @, #, RT, emojis have been removed and the sentence have been tokenized.

```

# helper function to clean tweets
def processTweet(tweet):
    # Remove HTML special entities (e.g. &#x2013;)
    tweet = re.sub(r'&\w*;', '', tweet)
    # Convert @username to AT_USER
    tweet = re.sub('@[\^s]+', '', tweet)
    # Remove tickers
    tweet = re.sub(r'\$\w*', '', tweet)
    # To lowercase
    tweet = tweet.lower()
    # Remove hyperlinks
    tweet = re.sub(r'https?:\/\/.*\/\w*', '', tweet)
    # Remove hashtags
    tweet = re.sub(r'#\w*', '', tweet)
    # Remove Punctuation and split 's, 't, 've with a space for filter
    tweet = re.sub(r'[' + string.punctuation.replace('@', '') + ']+', ' ', tweet)
    # Remove words with 2 or fewer letters
    tweet = re.sub(r'\b\w{1,2}\b', '', tweet)
    # Remove whitespace (including new line characters)
    tweet = re.sub(r'\s\s+', ' ', tweet)
    # Remove single space remaining at the front of the tweet.
    tweet = tweet.lstrip(' ')
    # Remove characters beyond Basic Multilingual Plane (BMP) of Unicode:
    tweet = ''.join(c for c in tweet if c <= '\uFFFF')
    return tweet

```

The tweets are filtered by using the location as “Nepal” with the hashtags #COVID-19 and #coronavirus. The gathered data has been stored in CSV format, and fed to the Sentiment Analysis library, namely, Textblob. The following table shows the retrieved data from Twitter API.

The following table shows the retrieved data from Twitter API in different dates:

Date	No. of tweets for #coronavirus and #COVID-19
21 st May 2020	82
22 nd May 2020	69
23 rd May 2020	54
24 th May 2020	33

25 th May 2020	66
26 th May 2020	53
27 th May 2020	49
28 th May 2020	61
29 th May 2020	48
30 th May 2020	59
31 st May 2020	41
Total	615

After collecting the data to perform Sentiment analysis TextBlob library has been called. TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation and more[25]. The main focus on data collection is on the content and timestamps of the tweets.

Sentiment analysis has a certain procedure that begins with grabbing the collected data, then identifying the data. The final decision is conducted in the phase of sentiment polarity as well as subjectivity. According to the TextBlob documentation[25], TextBlob takes advantage of Naïve Bayes (NB) model for classification. NB classifier has been trained on NLTK (Natural Language ToolKit) to detect valence of aggregated tweets.

```
#code for sentiment analysis
from textblob import TextBlob
from textblob.sentiments import NaiveBayesAnalyzer

def getSubjectivity(text):
    # return TextBlob(text).sentiment.subjectivity
    blob = TextBlob(df['text'], analyzer=NaiveBayesAnalyzer())
    return blob.sentiment
```

NB is probabilistic algorithm that uses Bayes theorem to compute sentiment distribution over the data. However, NB dissects any text to a bag of words which means the positions of the words are completely disregarded. The Bayes equation to predict the sentiment probability is:

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(\text{features}|\text{label})}{P(\text{features})}$$

Where $P(\text{label})$ is the prior probability of a label, $P(\text{features}|\text{label})$ is the prior probability that a given feature set is being classified as a label, and $P(\text{features})$ is the prior probability that a given feature set is occurring.

After that, function has been created to classify the sentiment as “Positive”, “Negative” and “Neutral” as follows:

```
#create a function to compute the negative, neutral and positive analysis
def getAnalysis(score):
    if score < 0 :
        return 'Negative'
    elif score == 0:
        return 'Neutral'
    else:
        return 'Positive'
df['Analysis']=df['Polarity'].apply(getAnalysis)
df
```

The result shows the respective sentiments of the tweets:

	Unnamed: 0	id_str	from_user	text	created_at	user_location	Subjectivity	Polarity	Analysis
0	0	1.260000e+18	PradipGajurel4	over million children class home due the lockd...	5/21/2020	Kathmandu Nepal	0.502778	0.136111	Positive
1	1	1.260000e+18	thisweekinnepal	confinamiento causa del limpiado aire sobre no...	5/21/2020	Nepal	0.000000	0.000000	Neutral
2	2	1.260000e+18	thisweekinnepal	most people lie below poverty line they can ev...	5/21/2020	Nepal	0.225000	0.200000	Positive
3	3	1.260000e+18	Lalitydv45	these are the meals being served daily sano pa...	5/21/2020	Birganj	0.333333	0.000000	Neutral
4	4	1.260000e+18	ILO_Nepal	what can governments employers and workers hel...	5/21/2020	Nepal	0.533333	0.016667	Positive
...
610	611	1.270000e+18	DB_Remastered	nepal extends nationwide lockdown until june j...	5/31/2020	From Pokhara, Nepal	0.000000	0.000000	Neutral
611	612	1.270000e+18	thehimalayan	considering the increasing number coronavirus ...	5/31/2020	Kathmandu, Nepal	0.000000	0.000000	Neutral
612	613	1.270000e+18	katuwal_hemant	govt isadmitting its inability care for citize...	5/31/2020	Nepal	0.601429	-0.182857	Negative
613	614	1.270000e+18	FranceInNepal	confinement prolonge jusqu' juin\lockdown ext...	5/31/2020	Lazimpat, Kathmandu	0.100000	-0.050000	Negative
614	615	1.270000e+18	Ekendra	day good morning from windy and cooler 1446 ju...	5/31/2020	Nepal	0.600000	0.250000	Positive

615 rows × 10 columns

After the sentiment analysis, list of positive, negative and neutral texts can be shown as follows:

1)most people lie below poverty line they can even save money for future this present scenario daily wage workers are helpless they say its better die virus then because hunger
 3)what can governments employers and workers help fight they can work together social dialogue key
 7)"this incident like cover that they are doing hide their incompetency" test and cases control where stand wise don' let nationalism sweep you off divert from the main concern
 13)thank you for your kind support towards plha community nepal
 20)with new cases single day tally reaches 457

Fig: List of positive tweets

1)isn this exactly what deplorable indian media doing abhorrent
 6)impossible buy things like sanitary pads maintain our hygiene"-naajma khatun kapilvastu she among dozens women who while staying quarantine sites various parts nepal receiv
 14)seriously what does mean that the persons who died did not die but due other complications had report death due because who mandate being too technical here too ignorant
 23)personal essay family' loss nepal weak public healthcare and the stigma
 43)anyone who has ever been hospitalised undergone treatment one kathmandu' private hospitals knows just how expensive healthcare can get family' loss weak public healthcare

Fig: List of negative tweets

2)these are the meals being served daily sano paila patients once again like heartily thank our staff hospital crew local administrative bodies the police and all helping b
 5)nepal daily update 2020 lockdown day sources
 15)our sharing about nhn integrated healthcare model covid knowledge exchange titled community health workers and chronic disease the time covid
 29)doesn affect everyone equally million people are living with hiv and that makes them more vulnerable let not lose the progress have made our fight
 48)mcc' 2nd compact invests power infrastructure improve reliability benefiting key hospitals the country one these institutions the korle teaching hospital which treating
 75)video health ministry confirms third death the country
 104)nepal disaster risk reduction and management authority covid19 dashboard

Fig: List of neutral tweets

7. Results and Discussion

This section show the Positive versus Negative Versus Neutral for sentiment polarity and objective versus subjective versus Neutral for subjectivity results are shown and discussed.

Sentiment Polarity:

Following bar chart demonstrates sentiment polarity over Twitter for 11 consecutive days starting in 21st May 2020. The data has been shown by the combined keywords #COVID19 and #coronavirus. Units are measured in percentage. We know that total number of tweets are 615.

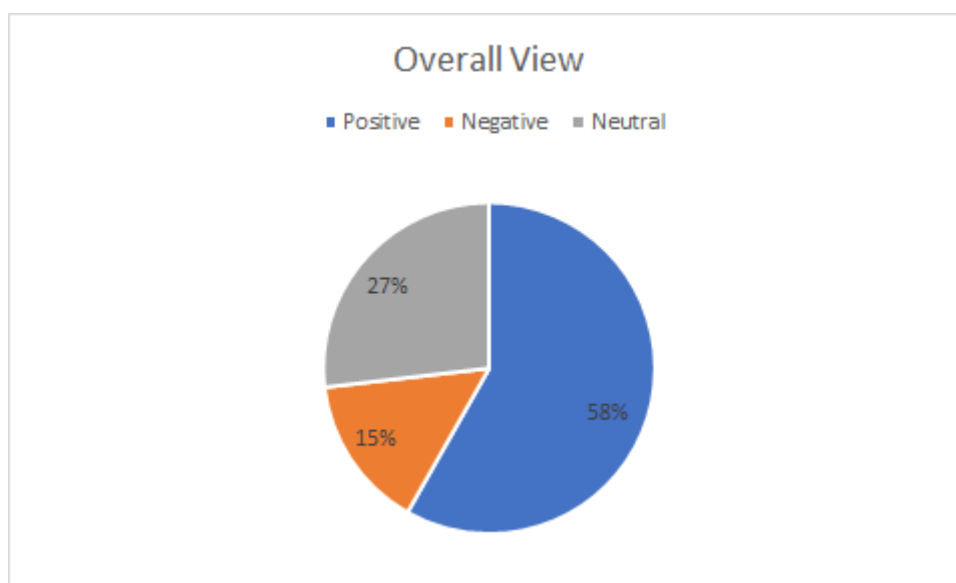
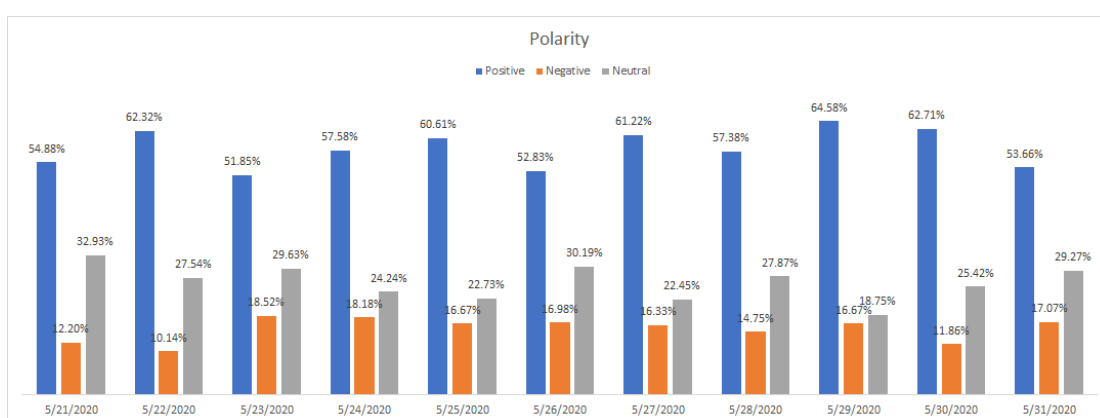


Fig: Overall view

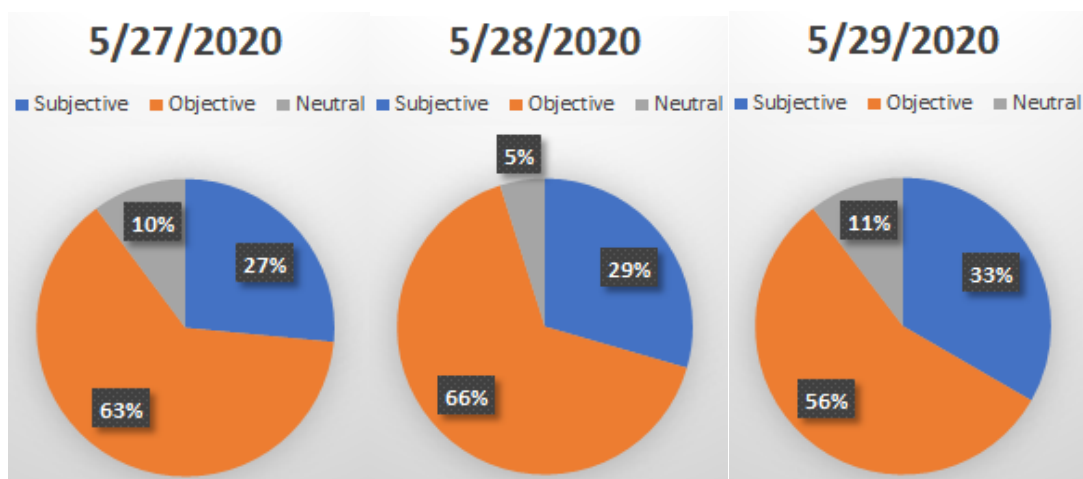
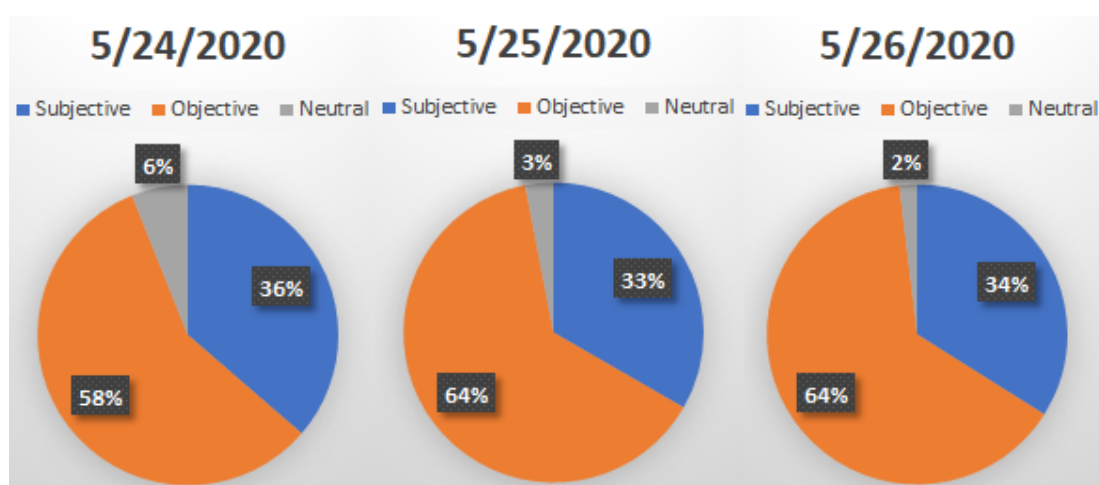
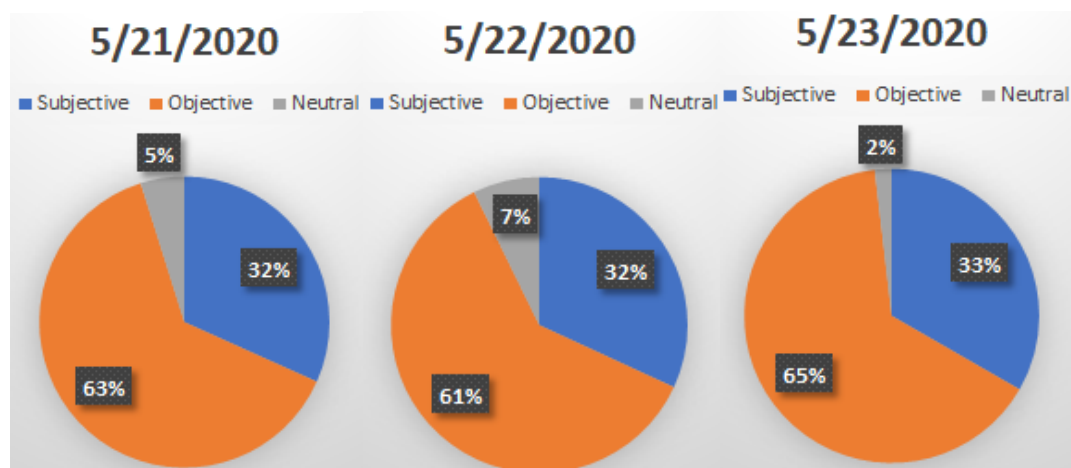
Overall 58% of people published optimistic views, while only around 15% of the tweets were negative. However, the neutral toll regarding both coronavirus and covid-19 keywords was 27%.



In 21st May, 54.88% tweets were positive and 12.20% are negative. It is found that 32.93% tweets were neutral. The most positive tweets can be observed on 29th May 2020 i.e. 64.58%. The figures of different days are relatively similar. Therefore, there is no significant sharp fluctuation in the tolls.

Subjectivity

In terms of the perspective of tweeters, there are subjective and objective viewpoints. There are eleven pie charts for combined #coronavirus as well as #COVID-19 keyword.



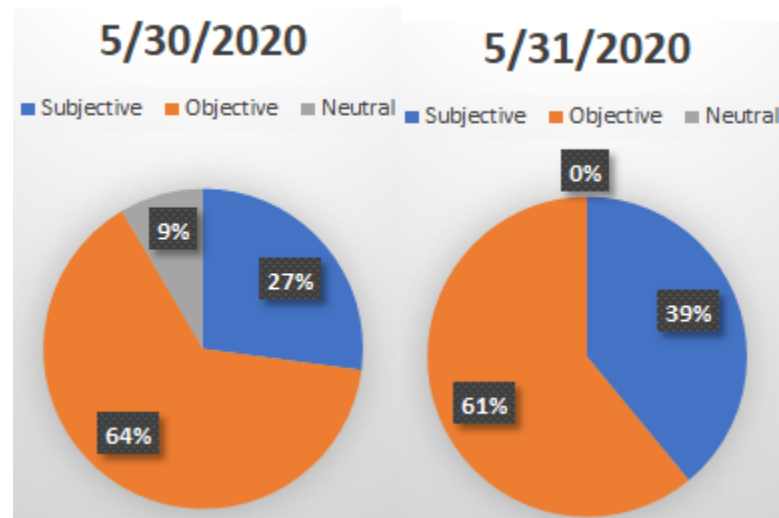


Fig: Subjectivity of tweets in different days

It can be observed that their pie chart portions are very similar, as a result of subtle variance data. Overall, the large portion of the records were objective which was in average 62.27%. Meanwhile it seems that in average 32.37% were being subjective of which expressed their feelings and opinions. Lastly, only in average 5.36% of the tweets have no clear characteristic to be neither subjective nor objective.

Discussion of sentiments

The aim of this analysis is to identify the emotional state of people about COVID19. This can be interpreted in the following chart:

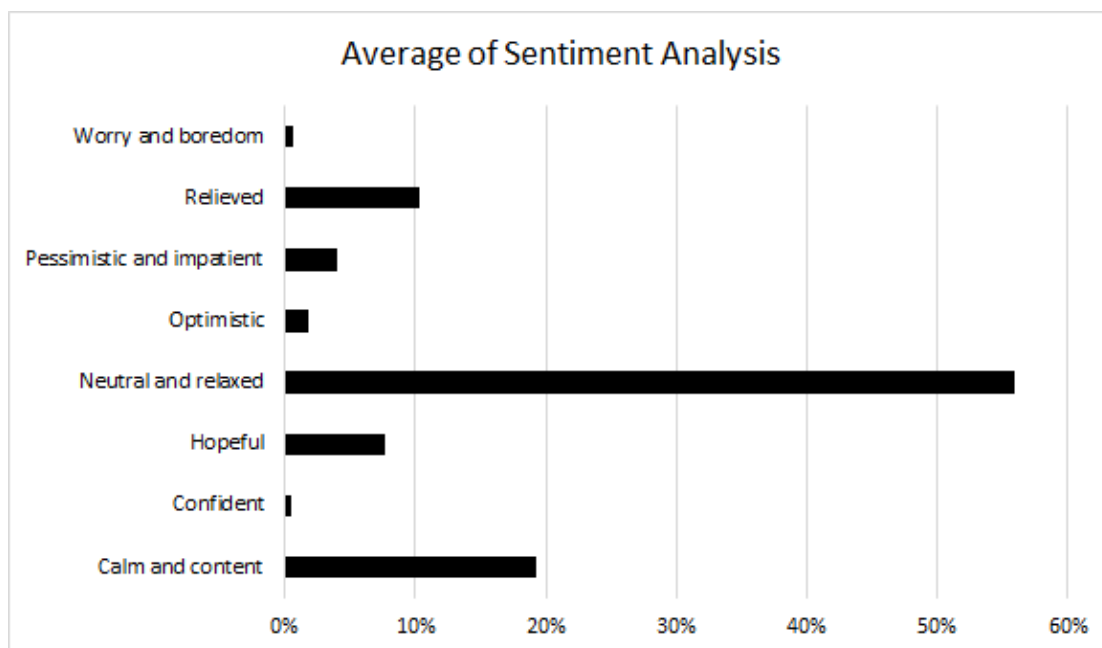


Fig: Average of sentiment analysis

From the above chart it is found that, the majority of the reactions i.e. 56%, toward were Neutral and relaxed. In addition, 8% and 19% of the tweets involved the feeling of hopeful and calm respectively. Negligible amount of tweets (only 3) found confident. Additionally, 2%, 4% and 10% are found Optimistic, Pessimistic and Relieved respectively. The minority only comprised 1% of the cluster found Worry and boredom.

In order to have a clear perception of the above figure, the following table shows emotion analyzing technique in this research. The below table has been applied as the “Emotional Guidance Scale” for polarity evaluation[22]. The higher emotion on the scale which is happy and joy and its value is 1 denotes better-feeling, while the lower on the scale which is and its value is -1 depressed and fear denotes more negative emotional states, in addition to the middle emotional state with 0 value represents neutral and relaxed state. Also, the rest of the feelings can be distributed evenly as spectrum between these three mentioned scales.

Table: Emotional Guidance Scale

Scale	Emotion
1	Happy and joy
0.8	Confident
0.6	Optimistic
0.4	Hopeful
0.2	Calm and content
0	Neutral and relaxed
-0.2	Relieved
-0.4	Pessimistic and impatient
-0.6	Worry and boredom
-0.8	Discouraged and difficulty
-1	Depressed and fear

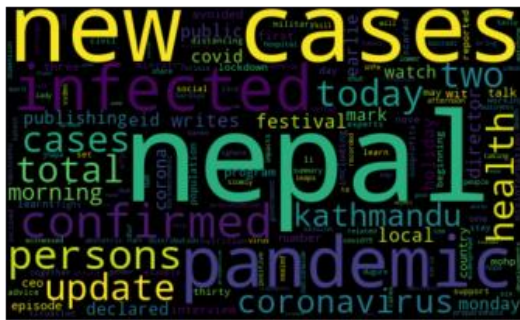
The daily wordcloud have been presented as follows:



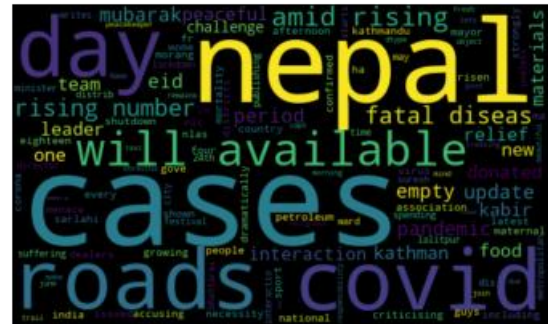
21st May 2020



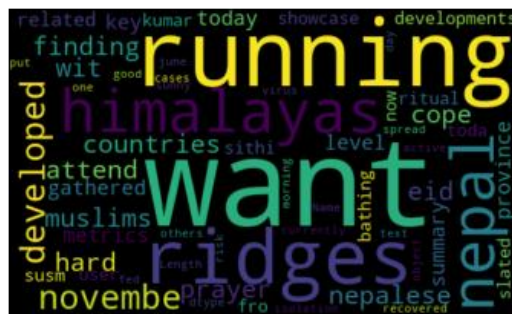
22nd May 2020



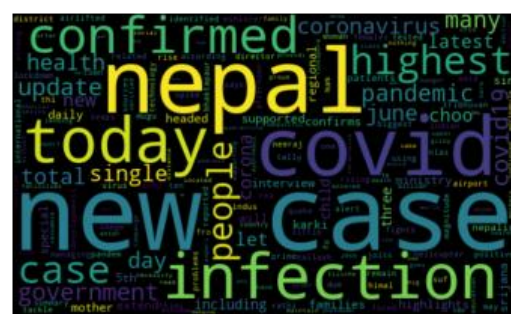
23rd May 2020



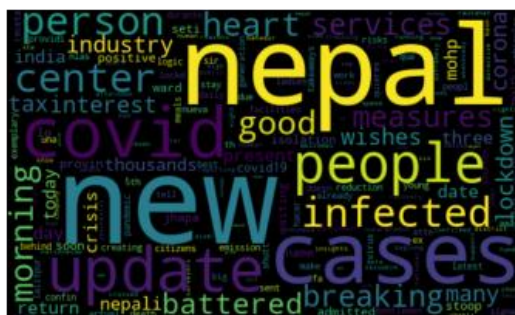
24th May 2020



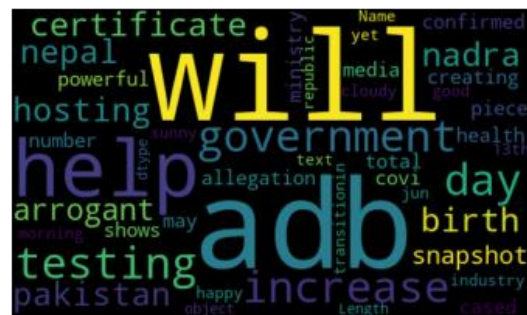
25th May 2020



26th May 2020



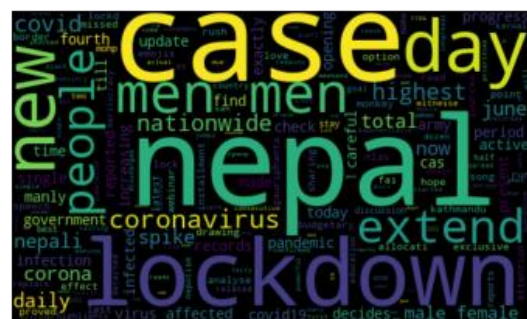
27th May 2020



28th May 2020



29th May 2020



30th May 2020

- Most of the user don't share their location so, only few user has mentioned their location as Nepal. Hence, it decreased the volume of tweets data.
- Twitter only allows to collect data from last 7 days in normal condition.

10. Future works

- In this paper Textblob library has been chosen which is using Naïve Bayes model. But there are other models that may provide interesting results such as lexicon-based algorithms.
- This research application is not only convenient for coronavirus health issue but it can also be adopted as model to discover sentiment emotion for the future similar cases.

References

- [1] A. D. Dubey, "Twitter Sentiment Analysis during COVID19 Outbreak," no. March, pp. 1–9, 2020.
- [2] "Health Sector Emergency Response Plan," no. May, 2020.
- [3] "Health Sector Response to COVID-19 Key Activities ;," pp. 1–3, 2020.
- [4] G. Shobana, B. Vigneshwara, and M. S. A, "Twitter Sentimental Analysis," no. November 2018, 2019.
- [5] V. K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, and S. Stehle, "Sentiment analysis during Hurricane Sandy in emergency response," *Int. J. Disaster Risk Reduct.*, vol. 21, no. May 2016, pp. 213–222, 2017.
- [6] M. Harvinder and J. Kaur, "Sentiment Analysis from Social Media in Crisis Situations," pp. 251–256, 2015.
- [7] D. Anagnostopoulos, E. Psomakelis, K. Tserpes, and T. Varvarigou, "COMPARING METHODS FOR TWITTER SENTIMENT," no. January, 2014.
- [8] C. Kaur and A. Sharma, "EasyChair Preprint Twitter Sentiment Analysis on Coronavirus using Textblob," 2020.
- [9] M. Dhanush, I. N. S, A. Patra, P. Biswas, and G. Immadi, "Sentiment Analysis of a Topic on Twitter using Tweepy," pp. 2–5, 2018.
- [10] A. Bastola *et al.*, "The first 2019 novel coronavirus case in Nepal," *Lancet Infect. Dis.*, vol. 20, no. 3, pp. 279–280, 2020.
- [11] R. Muthusami, A. Bharathi, and K. Saritha, "Covid-19 outbreak: Tweet based analysis and visualization towards the influence of coronavirus in the world," *Gedrag en Organ.*, vol. 33, no. 2, pp. 534–549, 2020.
- [12] M. O. Lwin *et al.*, "Global sentiments surrounding the COVID-19 pandemic on

- Twitter (Preprint),” *JMIR Public Heal. Surveill.*, vol. 6, pp. 1–4, 2020.
- [13] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, “COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification,” *SSRN Electron. J.*, no. May, pp. 1–21, 2020.
 - [14] N. K. Rajput, B. A. Grover, and V. K. Rathi, “Word frequency and sentiment analysis of twitter messages during Coronavirus pandemic,” no. April, 2020.
 - [15] H. Drias, T. Houari, and Y. Drias, “Mining Twitter data on COVID-19 for Sentiment analysis and frequent patterns discovery,” 2020.
 - [16] O. Gencoglu and M. Gruber, “Causal Modeling of Twitter Activity During COVID-19,” pp. 1–12, 2020.
 - [17] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. Lehmann, “An ‘ Infodemic ’: Leveraging High -Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak,” 2020.
 - [18] A. Abd-alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah, “Top Concerns of Tweeters During the COVID-19 Pandemic : Infoveillance Study,” vol. 22, pp. 1–9, 2020.
 - [19] Mohammed Alhajji, A. Al Khalifah, M. J. Aljubran, and M. Alkhalifah, “Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19,” no. April, 2020.
 - [20] R. P. Kaila and A. V. K. Prasad, “INFORMATIONAL FLOW ON TWITTER - CORONA VIRUS OUTBREAK – TOPIC MODELLING APPROACH,” *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 3, pp. 128–134, 2020.
 - [21] K. Suppala and N. Rao, “Sentiment analysis using naïve bayes classifier,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 8, pp. 264–269, 2019.
 - [22] K. H. Manguri and P. R. M. Amin, “Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks,” *Kurdistan J. Appl. Res.*

- [23] “Tweepy Documentation — tweepy 3.8.0 documentation.” [Online]. Available: <http://docs.tweepy.org/en/latest/index.html>. [Accessed: 08-Jun-2020].
- [24] “Google Colab Free GPU Tutorial - Deep Learning Turkey - Medium.” [Online]. Available: <https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d#:~:text=Google Colab is a free,TensorFlow%2C PyTorch%2C and OpenCV>. [Accessed: 08-Jun-2020].
- [25] S. Lorla, “TextBlob Documentation,” *TextBlob*, p. 1, 2020.