Foundation University Journal of Engineering and Applied Sciences

FUJEAS
Vol. 4, Issue 1, 2023.
DOI:10.33897/fujeas.v4i1.841

Research Article

Sentiment Analysis of COVID-19 Tweets using Neural Network in Pakistan

Syed Tayyab Ul Mazhar, Hasnain Khan, Shazia Usmani, Uzma Afzal *

Federal Urdu University of Arts, Science and Technology, Karachi, Pakistan

* **Corresponding author**: uzma.afzal@fuuast.edu.pk

Abstract:

Covid-19 is a respiratory infectious disease that spreads from person to person. World Health Organization declared it a pandemic in March 2020. Social distancing and lockdowns during the pandemic affected the lives, sentiments, and mental health of people. They are mentally, physically, and emotionally disturbed. It was observed that the verification of the very first Covid case increased the precariousness and fear in Pakistan because people had no information about the state policies. In this paper, Covid-related tweets from Pakistan for the year 2021 are collected. Pakistan faced two Covid waves in 2021. Dataset is preprocessed and a neural network-based model is applied to analyze the sentiment of the tweet. Sentiments of people are analyzed separately for each Covid wave, moreover, a comparative analysis is presented to discuss the change in people's sentiments with the passage of time. Results show a reduction in the number of tweets from third to fourth wave. Moreover, the %decrease in negative and positive sentiments are 6 and 1 respectively which show that by the passage of time, people started living with the environment and time healed the wounds of fears and uncertainty in their lives.

Keywords: Sentiment Analysis; Covid-19; Lockdown; Twitter; Neural Network; Pakistan.

Article Citation:
Mazhar et al. (2023). "Sentiment Analysis of COVID-19 Tweets using Neural Network in Pakistan". *Foundation University Journal of Engineering and Applied Sciences*
DOI:10.33897/fujeas.v4i1.841

Published by
Foundation University
Islamabad.
Web: https://fui.edu.pk/

1. Introduction

In 2019, a novel coronavirus (Covid-19) caused a respiratory illness in Wuhan, China. It is an infectious disease that spreads from a person's cough, sneeze, or breath. Patients can have mild, moderate, or serious symptoms [1]. The virus was confirmed to have reached Pakistan at the end of February 2020 [2]. One case from Karachi and one from Islamabad were reported. Till December 2022, 1.57M cases of Covid-19 have been reported in Pakistan, of which almost 98% of patients were recovered and only 2% of patients were unable to survive.

The number of cases increased rapidly at the start of April 2020 and caused the first Covid-19 wave in Pakistan [3]. The highest number of cases reported in a day was more than 6k. This wave lasted till the end of July 2020. The Second Covid-19 wave started in October 2020 and ended in February 2021. The highest number of cases reported in a day was almost 3.7k. After the end of the second wave, the third wave started immediately in March 2021 and lasted till the end of June 2021. The highest number of cases reported in a day was almost 6k. Till the mid of 2022, Pakistan faced six waves of Covid-19 spread of different intensities.

Social distancing and lockdowns during the pandemic affected the lives and mental health of people globally because many countries faced the same severity and intensity of the Covid spread. This topic also caught the attention of many researchers who carried out different research-based projects to analyze the different dynamics of the disease. In [4], authors use the data set of Covid-19 tweets to analyze the user sentiments. The effect of the Covid-19 pandemic on people's sentiments in different countries is also discussed in [5]. Covid-19 also introduces and boasts the concept of working from home. Sentiments of people who work from home are discussed and analyzed in [6]. Similarly, in [7], [8], and [9] tweets related to Covid-19 are mined to analyze the sentiments.

Covid-19 also affected the sentiments of people in Pakistan. Social distancing and lockdowns (full and partial) disturbed people mentally, physically, and emotionally. It was noticed that the confirmation of the first Covid case increased the uncertainty and fear in people because they had no idea about the government's actions and policies. They were not sure about their work, jobs, education, and daily social life. Moreover, the use of social media was also increased during the pandemic.

In this paper, we collected data of covid related tweets from Pakistan (five representative cities, i.e., Karachi, Lahore, Quetta, Peshawar, and Islamabad) for the year 2021 (Sadiq and Qureshi 2010) and applied a Neural Network (NN) based model to classify the tweet sentiment We also analyzed the change in sentiments and presented a comparative analysis to discuss the different aspects of the sentiments which changed by the passage of time. For this, we selected the data of the third and fourth waves as we already mentioned that in 2021 Pakistan faced the third and fourth Covid waves.

The primary contributions of the research are as follows:

- We analyze the sentiments of people in Pakistan by presenting neural network-based models to classify the tweets during the third and fourth wave of the Covid-19 pandemic.
- We also present a comparative study to discuss the change in people's sentiments that occurred with the passage of time.

Results show a reduction in the number of tweets from third to fourth wave. Moreover, the %decrease in negative and positive sentiments are 6% and 1% respectively which shows that by the passage of time, people started living with the environment and time healed the wounds of fears and uncertainty in their lives, which ultimately helped them in returning to the normal routine life.

The rest of the paper is organized as follows. The background of the study including the situation of Covid-19 in Pakistan, the social network forum Twitter, and basic building blocks of the neural network is presented in Section 2. Section 3 discusses the current state-of-the-art related work. Sections 4 and 5 explain the data collection process and experimental setup respectively. Section 6 presents the results in detail. Finally, our work is concluded in Section 7.

2. Background

In this section, we present the background study of the factors which are the crux of our study. A detailed overview of Covid-19 is presented along with the primary elements of sentiment analysis and neural net.

2.1. Covid-19

Covid-19 is an infectious disease caused by the coronavirus. It has been shown that most people who get infected by this virus have either mild or moderate symptoms and do not need special care for recovery. In serious cases, the infected person can have symptoms of difficulty in breathing, chest pain or pressure, and loss of speech or movement.

The virus spreads in the form of microscopic liquid particles from the mouth or nose of the infected person. Larger respiratory droplets to smaller aerosols are among the particles. If anyone is near the COVID-19 patient then there is a chance of getting the infection by breathing in the same air, or by contacting a contaminated surface and then touching his eyes, nose, or mouth. The virus disseminates in close and congested places.

On 26 February 2020, the *Covid* virus was confirmed to enter Pakistan, a case from Karachi and a case from Islamabad were reported. At the time of writing this paper, there have been almost 1.57M confirmed cases, of which 1.54M have recovered, and 30k were unable to survive. Till now, Pakistan has faced four waves of different intensities. Fig. 1 shows an image of the Covid-19 waves along with the number of reported cases.

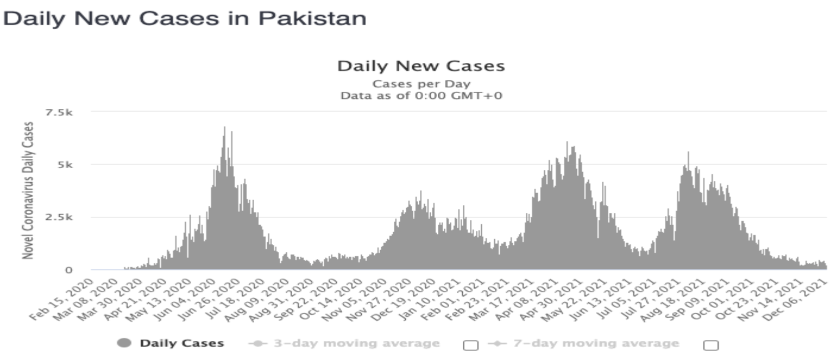

Figure 1: Covid-19 cases in Pakistan

2.2. Sentiment Analysis and Twitter

Sentiment Analysis (SA) is a technique of NLP (Natural Language Processing) that identifies the emotion of the text. There are three different levels of categorizations of SA, i.e., document, aspect, and sentence. The objective of document-level is to analyze an opinion paper as expressing a positive-negative opinion. The whole document is viewed to be a unit of primary information (talking about one topic). Sentence-level SA categorizes the sentiment of every sentence by determining the subjectivity or objectivity of the sentence. It examines whether the sentence communicates positive or negative opinions. Because sentences are merely small documents, there is no primary distinction between document and sentence-level categories. Detailed opinions on all parts of the entity are not provided in the document and sentence-level SA. Aspect level SA is used to gain this detailed information. The main objective of aspect-level SA is to analyze sentiment in relation to particular features of entities. Fig. 2 shows a typical flow of SA working. It starts with reviewing the given text and keeping sentiment identification as an ultimate goal. Opinionative words are selected and classified. Finally, a polarity is assigned to the given text on the basis of the classification. Twitter is a social media platform that allows users to share their thoughts in the form of messages called tweets.

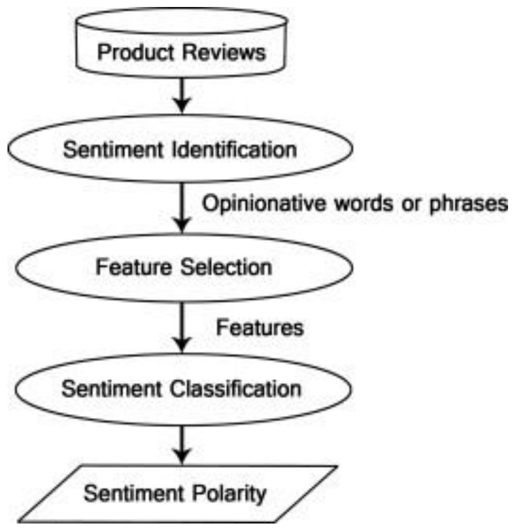

Figure 2: A typical flow of sentiment analysis

Twitter is one of the most useful and attractive domains for researchers due to the public availability of large amounts of tweet data on specific/general topics. For our research work, we used Twitter to collect data for sentiment analysis. Twitter is popular among researchers who work on SA. It provides APIs for collecting tweets, by date/time, and also allows filtering by language. We've used Twitter's API through a library called Tweepy. Tweepy is a popular library that is free and offers clean & easy to understand functions.

2.3. Artificial Neural Network

Artificial Neural Network (ANN) is a subset of machine learning and the heart of deep learning methods. ANN is inspired by human brain structure/function and mimics how organic neurons communicate with one another. Each neuron (node layer) has an output, an input, and one or many hidden layers. These nodes are connected to the others and have associated weights and thresholds. Node is activated and data is transferred to the next tier of the net if the value of a node output crosses a particular threshold value. Otherwise, no data is transferred to the next tier of the net.

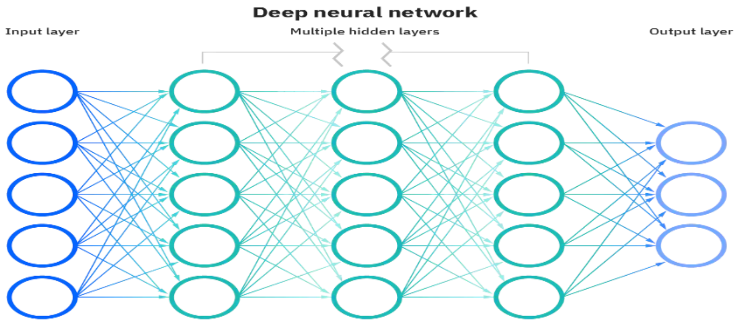

Figure 3. A deep neural network illustration

Neural networks use training data to learn and improve the accuracy of the model over time. Once the learning algorithm is fine-tuned for accuracy, it becomes a formidable tool in model implementation by allowing quick categorization and clustering of data. In the comparison of manual identification by an expert, objectives in speech and picture recognition can hardly take seconds/minutes rather than hours. One of the most well-known neural networks is the search algorithm of Google.

3. Related Work

In [4], the authors use the data set of covid-19 tweets to analyze the user's sentiments. R is used to build the model along with the descriptive textual analytics. Naive Bayes and regression classification models are also discussed to improve the proposed work. The effect of the Covid-19 pandemic on people's sentiments is discussed in [5]. Tweets from different countries are collected using the Twitter API and RTweet package. Data is cleansed by removing white space and stop words. NRC emotion lexicon is used on cleansed data to assign scores to the tweets' emotions. Results show that most of the people faced and dealt with this pandemic situation in a positive way. Researchers also highlight the fear and negative sentiments. Among the 8 selected countries, France, Switzerland, the Netherlands, and USA showed the negative sentiments.

In [6], the authors discuss the sentiments of people who work from home. LSTM, ANN, VADER, Naive Bayes, NRC Lexicon, Maximum Entropy, and MLP Classifiers are used to analyze the problem. Results show the trend of negative sentiments in people who do not physically attend their offices. Along with the negative trends, people of Bangladesh, Pakistan, Mali, and South Africa show positive trends too. People in Australia, India, Canada, the USA, Turkey, the UK, and Brazil show very negative sentiments. Using the NRC lexicon it is observed that Thailand, Vietnam, and Poland had the highest score for fear, whereas Oman, Syria, and Kazakhstan had high trust scores.

In [7], the Suzyhet Package of R is used to analyze the tweets' sentiments. The focus of the study is to review the emotional effect of tweets that were posted during the Covid-19 pandemic. A set of 10K tweets, retweets, and replies with #COVID19 hashtag is recorded. The TwitteR package is used to record the tweets. NRC sentiment lexicon is used to calculate the scores of positive/negative emotions. Results show that overall the tweets showed a positive sentiment. Among positive sentiments, joy and surprise had a low score, and among negative sentiments, fear had the highest score. Similarly, in [8] and [9] tweets related to Covid-19 are mined to analyze the sentiments. A typical SA process is followed to obtain the results. In [8], NodeXL is used to collect data Frequent patterns are mined using an adapted FP-growth algorithm. NRC lexicon is used to score the sentiments. In [9], authors find an effective machine learning-based model to obtain the optimal sentiment analysis prediction for coronavirus. This learning model is then employed in real time. To achieve better real-time performance, a system is developed in two components, i.e., an offline sentiment analysis component and an online prediction pipeline. Five ML Algorithms are used and compared to find the optimal online component model. Twitter Streaming API, Kafka, and Spark are used to develop the online prediction pipeline. Results validate that the Random Forest model using the unigram feature extraction method obtained the optimal performance which was finally employed to predict sentiment on Twitter streaming data for coronavirus.

In [10], Covid-19 related tweets are mined using Natural Language Processing. TextBlob, IBM Watson Tone Analyzer, BERT, and Mallet are used to automate the system. Results show mixed sentiments when the first case is reported, which shows that people were aware of the pandemic from the beginning. In [11], authors propose a model to forecast the change in stocks based on sentiments of headlines. The accuracy of the proposed model is 86.24% which validates the prediction results.

4. Data Collection and Preprocessing

We needed covid-19 related tweets to analyze the sentiments of the people of Pakistan. Unluckily, no such dataset was available in any public data repository so we had to generate the dataset. We used Twitter's API *Tweepy*, a popular library that provides methods for a smooth data collection process.

We collected tweets related to Covid-19, from Pakistan on a daily basis. In order to collect relevant tweets we used a set of keywords based on covid such as covid, virus, and lockdown. A tweet was fetched and collected if its text matched with any of the search keywords. A status object is returned by API for each fetched tweet that contains irrelevant information (in the context of our objective) along with the relevant information. To figure out this issue we created a separate Tweet class as shown in Fig. 4.

```python
class Tweet:
    def __init__(self, tweet_id, text, tags, created_at):
        self.tweet_id = tweet_id
        self.text = text
        self.tags = tags
        self.created_at = created_at

    def __str__(self):
        return f'Tweet ID:{self.tweet_id}  Tags:{self.tags}  Created At:{self.created_at}  Text:{self.text}'
```

Figure 4: A tweet class

The listing of the *Tweet class* shows that we stored information which were required to fulfill our research objectives such as tweet id, tweet text, and date-created along with the tags. In order to fetch the maximum number of relevant tweets, we used pagination which is provided by *tweepy. Cursor* method. As mentioned earlier, for each single tweet the API returns a status object and this status object does not contain the full text of the fetched tweet. Although we needed the full tweet text to analyze its sentiment accurately so we used tweepy.get_status method, which takes a tweet id as an argument and returns the complete text of the tweet (shown in Fig. 5). Then, we store the returned tweet into an array and finally these all returned tweets are saved in a CSV file.

Tweets usually contain noise, such as mentions, urls, punctuations, numbers, and special Characters. To remove the noise we used regex. We also removed short words with a character length of ≤ 2 (Fig. 6). Next, we tokenized each tweet and split each tweet into words to perform stemming on those words. Stemming is a process where each word is transformed into its root form. We also removed stop words from the tweets.

```
def processStatus(tweetID)->Tweet:
    status = api.get_status(tweetID, tweet_mode="extended")

    text = extractFullText(status)
    tags = extractTags(status)
    created_at = extractDate(status)

    tweet = Tweet(tweet_id=tweetID, text=text, tags=tags, created_at=created_at)

    return tweet
```

Figure 5: Fetching the complete text of a specific tweet

```
[ ] def clean_data(df, colName):
      # remove @user
      df['Tweet'] = np.vectorize(remove_pattern)(df[colName], '@[\w]*')
      # remove urls
      df['Tweet'] = df['Tweet'].apply(lambda x: re.split('https:\/\/.*', str(x))[0])
      # remove Punctuations, Numbers, and Special Characters
      df['Tweet'] = df['Tweet'].str.replace('[^a-zA-Z#]+',' ')
      # Removing Short Words
      df['Tweet'] = df['Tweet'].apply(lambda x: ' '.join([w for w in x.split() if len(w) > 2]))
```

Figure 6: Noise removal from tweets

	A	B	C	D
1	ID	Tags	Created At	Text
2	1398791054866911236		Sat May 29 23:59:10 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
3	1398790161954127876		Sat May 29 23:55:37 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
4	1398790136431792129		Sat May 29 23:55:31 +0000 2021	One of the greatest strengths of the covid core team is @fslsltn . Faisal Sultan
5	1398790062763020295		Sat May 29 23:55:13 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
6	1398789790783447047		Sat May 29 23:54:09 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
7	1398789033862574086		Sat May 29 23:51:08 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
8	1398788933698326531		Sat May 29 23:50:44 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
9	1398788912542277633		Sat May 29 23:50:39 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
10	1398788818128408579		Sat May 29 23:50:17 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
11	1398788725409292288		Sat May 29 23:49:55 +0000 2021	while it's good to recognize the work done & those who contributed, we m
12	1398788144355491848		Sat May 29 23:47:36 +0000 2021	We at the NCOC are committed to continue relentless efforts to overcome the
13	1398787968274468865		Sat May 29 23:46:54 +0000 2021	One of the greatest strengths of the covid core team is @fslsltn . Faisal Sultan
14	1398787750711726084		Sat May 29 23:46:02 +0000 2021	Alhamdulillah in the biggest challenge world has faced in last 100 years, Pakis
15	1398787368610635783		Sat May 29 23:44:31 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
16	1398787114544816132		Sat May 29 23:43:30 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
17	1398786820163387393		Sat May 29 23:42:20 +0000 2021	Alhamdulillah in the biggest challenge world has faced in last 100 years, Pakis
18	1398786607214477313		Sat May 29 23:41:30 +0000 2021	Alhamdulillah in the biggest challenge world has faced in last 100 years, Pakis
19	1398786530160807940		Sat May 29 23:41:11 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
20	1398786474536017920		Sat May 29 23:40:58 +0000 2021	Alhamdulillah in the biggest challenge world has faced in last 100 years, Pakis
21	1398786422920912896		Sat May 29 23:40:46 +0000 2021	Number of covid cases increasing every single day, can @ImranKhanPTI give
22	1398786391702740996		Sat May 29 23:40:38 +0000 2021	One of the greatest strengths of the covid core team is @fslsltn . Faisal Sultan

Figure 7: An image from the data set-cleaned

After the noise smoothing, we got the clean data set which was ready to pass the training model. Fig. 7 shows an image from the data set file. Neural networks work only on numerical data, whereas we have text data for tweets. So, we used vectorization to convert text to sequences of numbers.

As we mentioned we decided to collect data on a daily basis which is a very tedious task if it is done manually. So, we automated the data collection to ease the recording process. Manually running the script on a daily basis would be impractical for obvious reasons. So, we wrote a script that ran daily on the system which stored data. A new issue prompted here because we had to keep the system on and any fluctuations in recording could destroy the dataset. To solve the issue, we used an online tool called Deepnote [12]. Deepnote allows you to schedule a script, which means the script would run on a daily basis at the specified time. We scheduled our tweet collection script to run on a daily basis and all the repetitive tasks are automatically handled. We combined the daily collected data in two files, i.e., third-wave and fourth-wave covid data files.

5. Experimental Setup

5.1. Pre-labeled data Preparation for Training

Supervised learning requires pre-labeled data to train the prediction model [13]. For this purpose, we used the five different levels of sentiments, i.e., 'Neutral', 'Positive', 'Negative', 'Extremely Positive', and 'Extremely Negative'. The model we used generates the numeric outcome, so we replaced each sentiment level with a number. The new sentiment values range between 0 and 4. In the next step, pre-labeled data is split into training and testing datasets. The training set is used to train the model and the testing set is used to evaluate the trained model [14].

5.2. Neural Net Model Configuration

In this section, we explain the configuration settings we applied to tune our neural network-based prediction model. The hyperparameters of the model must be fine-tuned to obtain an accurate model [15]. To tune them optimally, we employed the trial and error method, which is typically used to tune the NN model parameters. In preprocessing, we transformed the array of text into 2D numeric arrays. That's why in the first layer of our NN model we applied an embedding layer. The embedding layer is a dictionary that links integer indices to dense vectors. This layer takes 2D integer tensors and two arguments, i.e., the number of possible tokens and dimensionality of the embedding (we tuned these parameters with 5000 and 20 values, respectively).

For the next layer, we used LSTM which processes sequences by iterating the sequence elements and keeping the relevant information [16]. In the LSTM layer, we proposed 15 hidden units within the layer with a 0.5 dropout rate. Dropout is a regularization technique for NNs to make sure the network does not depend totally on all its neurons; instead, enables itself to find more meaningful patterns in the dataset to increase the metric optimization. The last layer of the NN model was a dense layer with 5 output units to represent the 5 possible outcomes: 'Neutral', 'Positive', 'Extremely Negative', 'Negative', and 'Extremely Positive'. In order to generate probabilistic outputs, we used softmax as an activation function in this final layer.

5.3. Compiling the proposed NN model

After configuring and setting the hyperparameters of the NN model we compiled it by using RMSprop optimizer with its default learning rate. The optimizer continuously calculates the gradient of the loss and finds how to move against the loss function in order to find its global minima and discover the best network's parameters. Since the problem is a multiclass classification problem we used categorical_crossentropy as a loss function. The loss function evaluates how well the model works. If the predictions are good, it outputs a low number and if the predictions are off, it outputs a high number. Generally, it is a way to measure the goodness of the model. Specifically, it is a function to compute the distance between the current output of the algorithm and the expected output.

We also used checkpoints to save the best model during the training process. Checkpoints make sure that the model with the best accuracy is saved instead of the last model.

6. Results and Discussion

In this section, we present the results of NN based prediction model which we used to predict the sentiment of the given tweet. Table 1 shows the optimal values of the hyperparameters which are used to fine-tune the final prediction model. The model generated the optimal results with 1 hidden layer in 39 epochs with 0.001 value of learning rate and 32 batch size.

The accuracy and loss of the model with the validation dataset are shown in Table 2, i.e., 0.7258 and 0.7543 respectively. After tuning the hyperparameters, the LSTM model is trained with the optimal values of hyperparameters. The accuracy and loss of the trained model are 0.7767 and 0.6207 respectively (Table 2). Results show that the model generated better results during training.

Fig. 8 shows the results of sentiment analysis of Covid-19 tweets during the third wave of virus spread. 21% of tweets represent neutral sentiment while 18% and 14% of tweets show extreme emotions positive

Table 1: Optimal values for hyperparameters

Hyper Parameter	Optimal Value
num_layers	3
Epochs	39
learning rate	0.001
batch size	32

Table 2: Accuracy and loss of the model

	Accuracy	Loss
Training	0.7767	0.6207
Validation	0.7258	0.7543.

and negative respectively. In total, almost 45% (116K) of tweets show positivity and 34 % (89K) of tweets show the negativity of Pakistanis.

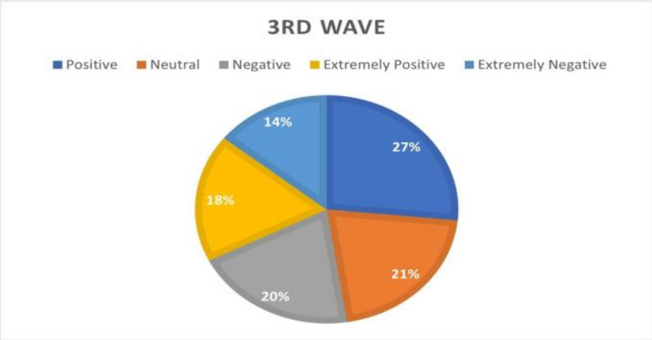

Figure 8: Sentiment analysis of Covid-19 tweets - Third-wave

Similar to Fig. 8, Fig. 9 presents the results of sentiment analysis of Covid-19 tweets during the fourth wave of coronavirus in Pakistan. 28% of tweets show a neutral sentiment while 16% and 10% of tweets show extreme emotions of positivity and negativity respectively. In total, almost 44% (29K) of tweets show positivity and 28% (14K) of tweets show the negativity of Pakistanis. It is also noted that the number of total tweets discussing Covid-19 was drastically reduced from the third wave to the fourth wave.

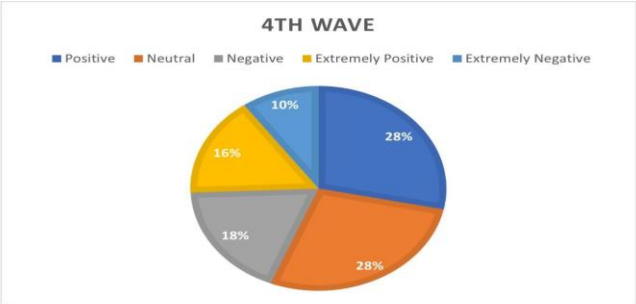

Figure 9: Sentiment analysis of Covid-19 tweets - Fourth wave

Fig. 10 presents a comparison of the sentiment analysis results for the third and fourth wave of virus spread, it also highlights the reduction in the number of tweets from the third to fourth wave which is almost 79%. This decrease in posting Covid-19 related tweets shows that with the passage of time, people started living with the environment, and the virus spread left affecting people's lives. Moreover, the %decrease in negative and positive sentiments are 6% and 1% respectively which also strengthens the argument that, with the time Covid-19 left hurting the sentiment of people in Pakistan and they returned to their daily routine life.

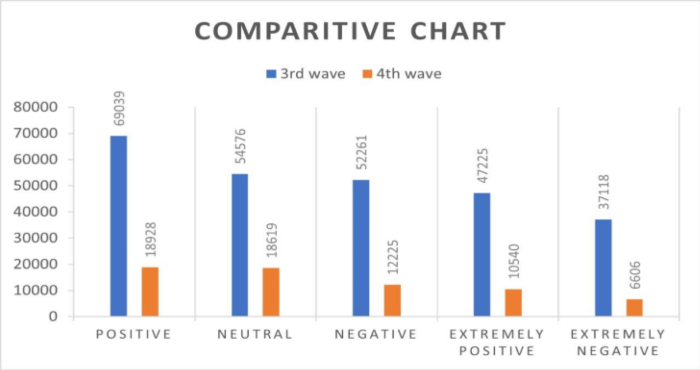

Figure 10: Sentiment analysis- a comparison

7. Conclusions

In this research work, Covid-related tweets from Pakistan for the year 2021 are collected to study the effects of social distancing and lockdowns on people's sentiments. These new social norms disturbed people's sentiments mentally and emotionally. A neural network-based model is employed to interpret and classify the tweet sentiment posted during the pandemic. The study revealed that the number of tweets are reduced from the third to fourth wave in 2021. Moreover, the %decrease in negative sentiments is higher than positive, i.e., 6%:1%. Results validated that by the passage of time, people started to opt for the current living environment which reduced fears and uncertainty in their lives. In the future, we will employ different machine learning techniques to analyze the sentiments and other Covid-related dynamics such as job uncertainty and economic crisis.

8. References

[1] G. Eason, B. Noble and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 247(935), pp. 529–551, 1955.

[2] N. Noreen, S. Dil, S. Niazi, I. Naveed, N. Khan, F. Khan and D. Kumar, "COVID 19 pandemic & Pakistan; limitations and gaps," *Global Biosecurity*, vol. 1(3), pp. 1–11, 2020.

[3] M. Shafi, J. Liu and W. Ren, "Impact of COVID-19 pandemic on micro, small, and medium-sized Enterprises operating in Pakistan," *Research in Globalization*, vol. 2, p.100018, 2020.

[4] J. Samuel, G. M. N. Ali, M. M. Rahman, E. Esawi and Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification," *Information*, vol. 11(6), p.314, 2020.

[5] A. D. Dubey, "Twitter sentiment analysis during COVID-19 outbreak," *Available at SSRN 3572023*, 2020.

[6] M. Mansoor, K. Gurumurthy V. R. Prasad, "Global sentiment analysis of COVID-19 tweets over time," *arXiv preprint arXiv:2010.14234*., 2020.

[7] A. Kumar, S. U. Khan and A. Kalra, "COVID-19 pandemic: A sentiment analysis: A short review of the emotional effects produced by social media posts during this global crisis," *European Heart Journal*, vol. 41(39), pp. 3782–3783, 2020.

[8] H. Drias and Y. Drias, "Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery," *medRxiv*, 2020.

[9] X. Zhang, H. Saleh, E. M. Younis, R. Sahal and A. A. Ali, "Predicting coronavirus pandemic in real-time using machine learning and big data streaming system," *Complexity*, pp. 1–10, 2020.

[10] W. Yu, N. Luo, X. Zhang and A. Li, "Twitter Sentiment Analysis Based on News Topics during COVID-19," [Online: Accessed: 2022] URL: https://towardsdatascience.com/twitter-sentiment-analysis-based-on-news-topics-during-covid-19-c3d738005b55

[11] Y. Jain and V. Tirth, "Sentiment analysis of tweets and texts using python on stocks and COVID-19," *International Journal of Computational Intelligence Research*, vol. 16(2), pp. 87–104, 2020.

[12] "Deepnote: Analytics and data science notebook for teams," [Online: Accessed: 2022] URL: https://deepnote.com/

[13] K. Das R. N. Behera, "A survey on machine learning: concept, algorithms and applications," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5(2), pp. 1301–1309, 2017.

[14] H. Jia, J. Liu, M. Zhang, X. He and W. Sun, "Network intrusion detection based on IE-DBN model," *Computer Communications*, vol. 178, pp. 131–140, 2021.

[15] U. Afzal, T. Mahmood, A. H. Khan, S. Jan, R. U. Rasool, A. M. Qamar and R. U. Khan, "Feature selection optimization in software product lines," *IEEE Access*, vol. 8, pp. 160231–160250, 2020.

[16] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228(10), pp. 2313–2324, 2019.