

Analysis Of Bangladeshi People's Emotion During Covid-19 In Social Media Using Deep Learning

Md. Sabbir Alam Pran
Dept. name of CSE
 Daffodil International University
 Dhaka, Bangladesh
 sabbir15-9267@diu.edu.bd

Md. Rafiuzzaman Bhuiyan
Dept. name of CSE
 Daffodil International University
 Dhaka, Bangladesh
 rafiuzzaman15-9655@diu.edu.bd

Syed Akhter Hossain
Dept. of CSE
 Daffodil International University
 Dhaka, Bangladesh
 aktarhossain@daffodilvarsity.edu.bd

Sheikh Abujar
Dept. of CSE
 Daffodil International University
 Dhaka, Bangladesh
 sheikh.cse@diu.edu.bd

Abstract—World is passing through a very uncertain circumstance as Coronavirus becoming a great threat. Staying in home is the best solution now to be safe. People are now passing their most of the time in social platform. They're reacting in public posts, news, articles and also commenting there. And a persons comment can talk about his sentiment. Emotion exploration is a very famous topic in the field of data mining. Lots of work have been done yet. In this piece of research, Bangladeshi people's comments on several Facebook news post related to coronavirus have been analyzed to observe the sentiment of them toward this situation. Using three classes investigation have been done on their emotions, which are Analytical, Depressed, Angry. The data set was developed in Bangla language. Several deep learning algorithms have been applied and found the maximum accuracy in CNN 97.24% and in LSTM 95.33%. Result shows that most people commented analytically. The outcome draw up the public psychology of Bangladesh toward the pandemic.

Index Terms—covid-19, sentiment analysis, social media, Bangladesh, epidemic, coronavirus, deep learning, facebook.

I. INTRODUCTION

More than 188 countries and territories have already been infected by coronavirus. About more than 365000 people died and this is the global situation up to 30 May 2020 [source: Johns Hopkins University] [1]. China Government and WHO (World Health Organization) already invented that this virus is the member of previous pandemic SARS (Severe Acute Respiratory Syndrome). People's life becoming very uncertain day by day as there is no curable vaccine. World's top researchers are working relentlessly to invent the vaccine. And till then the safety measurement given by WHO is the best way to survive from this deadly virus. All the affected countries are trying their best to survive by locking down the affected area. Using the facility of internet people are doing their work through online depending on their feasibility. The online activity of people is raising. Through social media the news pages are making post on the live situation regularly. People give their reactions, comments there. Their feedback through comment is a great reflection of their sentiment.

In 7th March Institute of Epidemiology, Disease Control and Research (IEDCR) discovered the first 3 affected cases in Bangladesh [2]. Bangladesh Government first declared country lock-down at march 26 [3]. Till 30th May 2020, 44608 total affected and 610 total death confirmed by IEDCR [4]. Almost all District of Bangladesh have been affected already [5]. Due to having a large number of Daily earners, the situation forced the Government to stabilize the lock-down situation from 31 May 2020 [6]. Bangladeshi people are very emotional by born. In last decade the number of internet users also increased in Bangladesh. Journalists are doing their best to give media coverage of every single events. In Facebook, lots of local media posting the news article about covid19. People share their opinion through comment section. We've collected lots of Bangla comment from several post related to covid19. Those posts contain the day by day affected and death news, salvation related news, funeral process, lockdown extension, treatment facilities, act of police during this situation, sacrifice of doctors and so on. People gave lots of potential feedback there.

In this study, we've tried to analyze the emotions or sentiment of Bangladeshi people in this crisis situations using Deep learning. We've analyzed 1120 data under three classes. Since CNN and LSTM did best in term of accuracy, we've implemented those model. This research will help every individual to understand their emotions and also help to take necessary steps to make themselves positive in this situation of pandemic.

II. LITERATURE REVIEW

Sentiment analysis is a computational bridge to people's inner opinions or emotions [7]. It's a very common area now to predict or analyze public's sentiment in different aspects such as review of customer on product [8], prediction on stock exchange [9], [10], news, blogs [11], political sentiment analysis through political articles, speeches [12] and so on.

Lots of data mining algorithms are using now-a-days to analyse and predict sentiment on several topic. By using data mining algorithms, people's sentiment can be judge through social media. Lot's of work have already been done on this area using data mining. LSTM got appreciable accuracy in analyzing public's opinion on Reddit [13]. Nave Bayes and KNN also got much better accuracy on tweet data to detect influenza in Arab region through people's social media sentiment [14]. We've used CNN for its better accuracy in our data set which is in Bangla language. However, tweeter is very common choice for researchers for easier way to analysis sentiment [15]. We've collected data from Facebook as it's the most well-known social media in Bangladesh. We've used three classes to define public's sentiment and those are Analytical, depressed, Angry. In [16] used two classes those are positive and negative. Where the topic on social-distancing got 71 percent positive opinion and work-from-home got 77.7 percent accuracy. In [17] used polarity. They've marked polarity as negative when the value goes under and not equal with respect to zero. For positive it's higher and not equal to zero and neutral for being equal. In this situation of pandemic covid19 social distancing is a very important sentiment. The more people maintain this the more they will be safe. In [18] said around 82.5 percent of tweet was in favor of social distancing. In social media comments, people's sentiment, emotion, reaction depends on the type of that comment. Controversial comment shows anger sentiment whether non-controversial comment express anxiety, sadness and more analytical thinking [19]. However due to become e lower income country, Bangladesh is in a very typical situation now. People's idea about this virus is not satisfactory [20], [21]. Government is doing their best to control the situation. The best solution for Bangladesh is now to become locked down properly [22]. That's the practical situation. We'll discuss about people's mental situation in Facebook. Bangladeshi people's opinion can be understood through this paper. Actually, we couldn't learn enough from past two prior pandemic SARS and MARS [23].

We barely found any work using Bangla data-set on emotion analysis. There are lots of emotion analysis related work with Twitter, Reddit except Facebook comment. As the the fight is with the invisible enemy, we should be positive as much as we can. Authorities should know peoples inner condition so that Authority can motivate them. So we've tried to analyse the mass emotion in this crisis. In our study, both CNN and LSTM used to train our model. We have also compared the both model for better understanding.

III. METHODOLOGY

Here in this part we will discuss about the whole working process and also the model which we've been used. Tensorflow 2.2.0 have been used to train and evaluate our model. Before train the model, we need to evaluate some pre-processing stuff for our data set. Which have been discussed in pre-processing

part. Then we'll introduce our applied model. Figure-1 shows the workflow structure and others described as below.

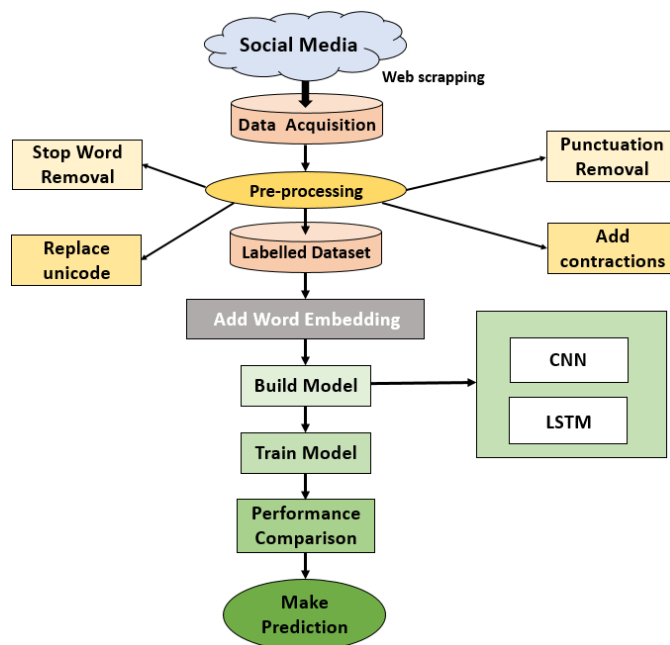


Fig. 1. Workflow.

A. Data Acquisition

Data is the heart of Artificial Intelligence. Machines efficiency depends on the quality and the amount of data. The more we can provide data the more machine can predict and analyze perfectly. In this study, we've collected data from Facebook which is in Bangla language. We've collected 1120 comments from several coronavirus related post. We've labeled those data in three classes based on comment type. Those are Analytical, Depressed and Angry. Analytical comment contains advice or suggestive type of speeches. Angry

Comment	Category/Class
এটি একটি খুব বিপজ্জনক ভাইরাস। আমাদের এই মারাত্মক ভাইরাস থেকে রক্ষা করার জন্য আমাদের সামাজিক দূরত্ব বজায় রাখা উচিত।	Analytical
অসন্তোষজনক। সামাজিক দূরত্ব পালনের ক্ষেত্রে যথাযথ হয়নি।	Angry
অনেক কষ্টে আছি। সারাদিন বাসায় বসে থাকতে হয়। কারো সাথে দেখা হচ্ছে না। আরো বড় বিষয় কারো বাসায় গেলে বাজে ভাবে নেয়। কিছু করার নাই। সবাই নিরুপায়। আমিও ভীত	Depressed

Fig. 2. Data Sampling.

and Depressed type of comment have been selected by the mood of those speeches. We've classified those comments semantically and those have been verified by two persons. We've finally got two columns. Those are category and opinion

as class/label. Figure-2 is an overview of sampling data. Among all of those comments we've found 394 under analytical, 363 for Angry and 363 in Depressed section. Figure-3 shows the pi-chart of all classes.

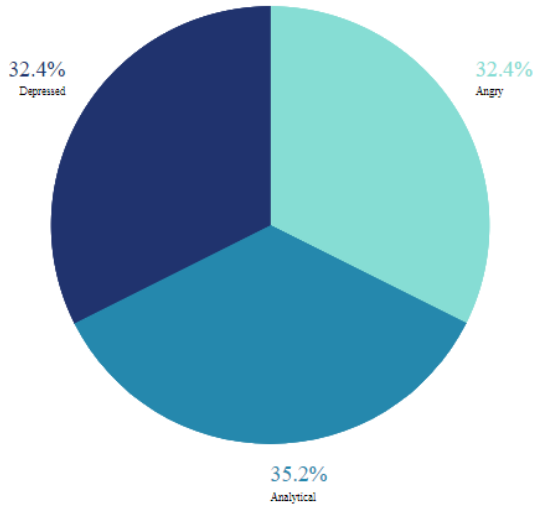


Fig. 3. Dataset.

B. Pre-processing

For getting proper output from a data analysis related work, data pre-processing has no alternative. Without doing this step misleading outcome will generate. A few steps have been followed to do this job. Figure-4 shows the structural view of those steps. Elaborate explanation is given below:

- **Add Contraction:** The shortened version of words or syllables are known as contractions. In English language, they mostly subsist in any one of written or spoken form. By dispelling specific letters and sounds, these concise versions or contractions of words are built. In case of English contractions, they are often created by dispelling one of the vowels from the word. For example, do not to don't and I would to I'd. In Bangla language those type of contractions have been added. Altering each contraction to its expanded, original form helps with text standardization.
- **Punctuation removal:** The data set had lots of unnecessary punctuations. Those have been removed technically.
- **Stop Word removal:** Stop word means commonly used word like "a","an","the" and so on. We've used Bangla data set. Which had lots of stop words. Those have been removed.
- **Replace Unicode:** It's mandatory to replace Unicode due to having the data set in Bangla.

C. Word Embedding

Word embedding gives us a path to use feasible, solid delegation in where same words get similar encoding. Moreover, this encoding does not need to do by hand. Embedding contains a dense vector which contains values in floating-point.

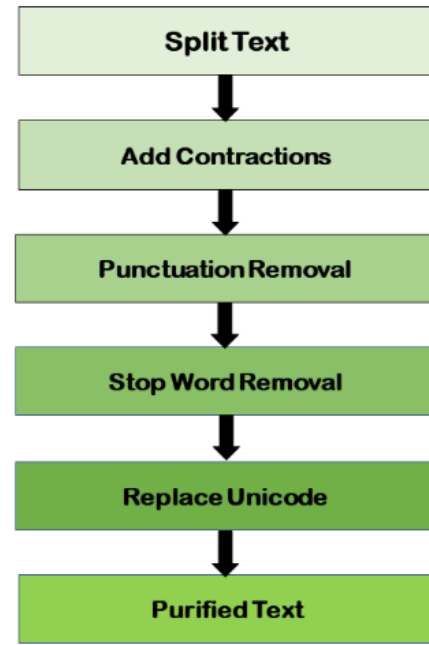


Fig. 4. Pre-processing steps.

Commonly embedding contain 8 dimensions. For large data set it can be up to 1024 dimensions. Embedding of higher dimension can draw up better relationship but need more data to learn. Words are represented by dense vector in an embedding. The jut of words within a continuous vector space describes by a vector. From text word's placement into the vector space is learned. It's done on the basis of the words which enclose the word during it's use. For our task we've used a pretrained word embedding "bnword2vec" that help us not to train our own embedding architecture and saves our time.

D. Applied Model

There are different types of models for working in data-driven techniques i.e. deep learning. These models do different types of work. For example, one type of model is used for image recognition and another type of model for working with speech data. In our work, We will use two different models i.e. (i) CNN and (ii) LSTM which is described below.

1) *Convolutional Neural Network:* CNN was first used in image classification tasks where convolution is the heart of images. Currently it is being used in NLP. [24]- [25] To make our work much simpler, we will describe to create our model on a single review. We've introduced a sequential model for our purpose. Firstly, input text that a review is represented as matrix that fed into an embedding layer which is used for lower dimensional representations. A pretrained word2vec of 300 dimesions and a vocabulary size of 4000 add into it. Then two Convolutional layers used for extract

features. We've used 64 and 128 filters with a 3x3 sized kernel, padding as 'valid' and stride as 1 into both convolutional layers to produce the feature map from the input. For adding non-linearity an activation known as ReLU used in both layers. For reducing dimensionality later a GlobalMax-Pooling layer used after both convolutional layers. Later a dropout of .25 have been introduced with a fully connected layer to reduce the overfitting of the model. Finally a softmax function used for distinguishing input review as three classes i.e. Analytical or Depressed or Angry. Fig-5 shows our applied CNN model.

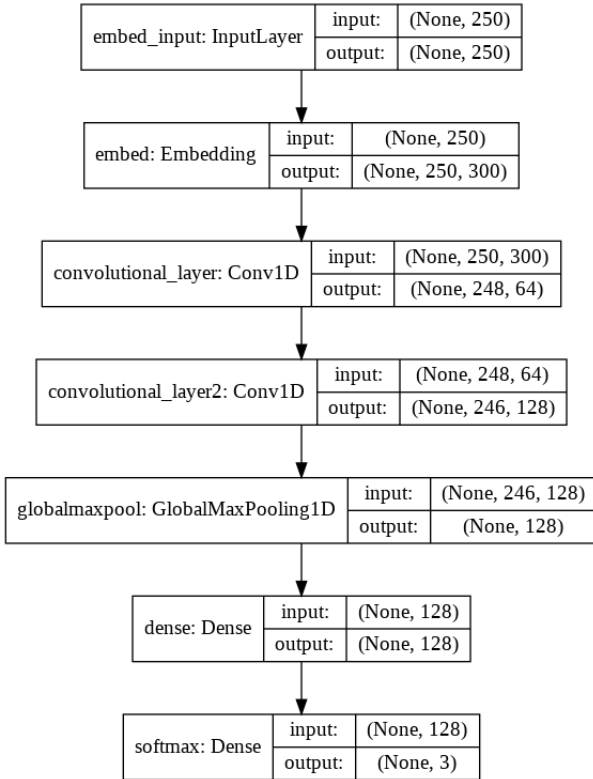


Fig. 5. Applied CNN.

2) *Long Short Term Memory*: A special type of neural network performs well in terms of textual data. It also performs really well in terms of long term dependencies. In the name of LSTM, there exist a term called memory. With this it's actually keeps previous information into its memory over a long time.

When it comes to use LSTM, the first few tasks of CNN will remain the same. Pretrained word2vec with 300 dimensions added into embedding layer with a vocabulary size of 4000. Then a ReLU activation used for adding non linearity. After that LSTM hidden layer of size 128 introduced. A dropout .25 used to reduce overfitting. Later a fully connected layer introduced to connect the output of all previous states. Finally a softmax function used for distinguish input review as three classes i.e. Analytical or Depressed or Angry. Fig-6 shows our applied LSTM model.

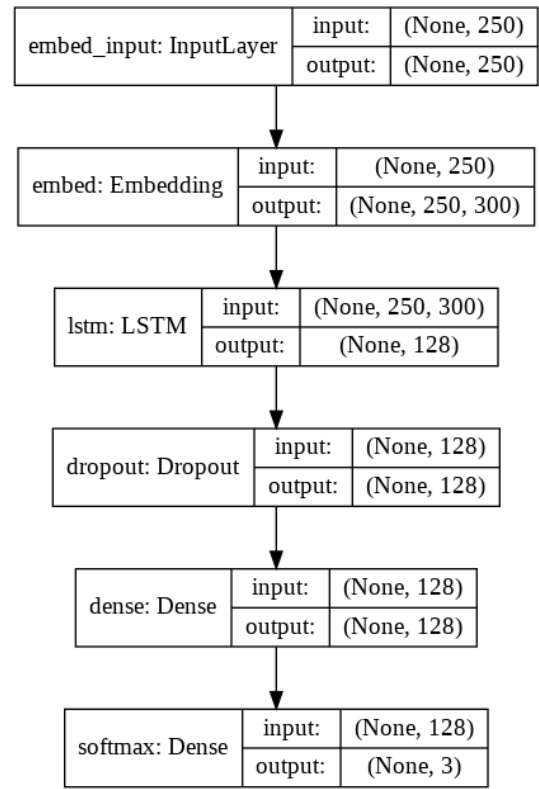


Fig. 6. Applied LSTM.

IV. EXPERIMENTAL RESULT & DISCUSSION

In this section, we have divided our experiment in two parts. For first of them we have introduced hyper-parameters of our experimented model. Then start training and got an accuracy. In the next part we will compare the performance of our two models and see which one works best.

A. CNN Experiment

First we set some hyper-parameters into our CNN model to measure the performance. Sometimes we have changed to check for which parameters our model gives better accuracy. Table-I shows the hyper-parameters that we've applied in our model.

TABLE I
CNN HYPERPARAMETERS

Hyper-parameter	Value
1st ConV size	64
2nd ConV size	128
Filter size	[3,3]
Batch size	128
Epoch	75
learning rate	0.001

Now we need to compile our model. Adam optimizer and categorical cross entropy loss function used to compile our model. Then we start training. After a successful training process we've got 97.24% accuracy.

B. LSTM Experiment

We set some hyper-parameters into our LSTM model to measure the performance. Sometimes we've changed to check for which parameters our model gives better accuracy. Table-II shows the hyper-parameters that we've applied in our model.

TABLE II
LSTM HYPERPARAMETERS

Hyper-parameter	Value
LSTM hidden state size	128
Dropout rate	.35
Batch size	128
Epoch	75
learning rate	0.001

Now we need to compile our model. Adam optimizer and categorical cross entropy loss function used to compile our model. Then we start training. After a successful training we've got 95.33% accuracy. Accuracy, loss curve and performance comparison introduced in next part.

C. Model Comparison

In this part the comparison of performance curves between those two model have been introduced and other evaluation metrics have discussed as follows.

Fig-7 & Fig-8 shows the accuracy and loss curves comparison where CNN performs really well and loss is also quite good.

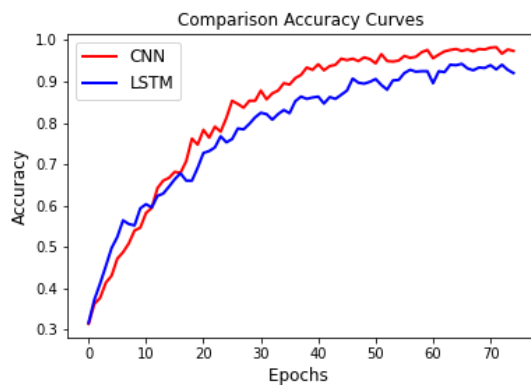


Fig. 7. Accuracy curve.

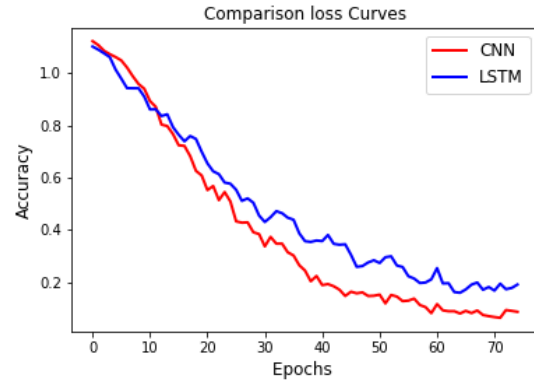


Fig. 8. Loss curve.

In Table-III average precision, recall and f1-score discussed. From that we can see both CNN and LSTM precision is 0.71 and 0.69 respectively whereas LSTM has .02 higher f1-score in compare to CNN.

TABLE III
AVERAGE PRECISION, RECALL AND F1-SCORE

Category	Precision	Recall	f1-score
CNN	0.71	0.72	0.70
LSTM	0.69	0.72	0.72

Table-IV shows the performance comparison of our applied model where CNN on top accuracy with 97.24%.

TABLE IV
PERFORMANCE COMPARISON

Model	Embedding size	Technique	Accuracy
CNN	300	word2vec	97.24%
LSTM	300	word2vec	95.33%

V. CONCLUSION AND FUTURE WORK

This is the very first research work on the sentiment toward covid-19 using Bangla language data set. We've got higher accuracy in CNN. And we've found that maximum comment is analytical types. Angry and depressed type of comments frequency is in second and third position. Maximum comments have been collected from Facebook news pages. The high frequency of analytical comments prove that Bangladeshi peoples are much positive toward this pandemic in social media. Since this is our first attempt, there are some limitations which can be improved. Due to insufficiency of Bangla data set and very different data frequency, precision has become very low. Diversity of data is also another cause of that.

Use of deeper layer is the best way to improve our model. Apply of other word embedding techniques can improve the precision. We will arrange more data and will use others word

embedding techniques along with ensure the use of deeper layer in future for betterment.

VI. ACKNOWLEDGEMENT

We gratefully acknowledge support from DIU NLP and Machine Learning Research LAB for providing GPUs support. We thank, Dept. of CSE, Daffodil International University for providing necessary supports. And also thanks to the anonymous reviewers for their valuable comments and feedback.

REFERENCES

- [1] Aljazeera.com. 2020. Coronavirus: Which Countries Have Confirmed Cases ?. <https://www.aljazeera.com/news/2020/01/countries-confirmed-cases-coronavirus-200125070959786.html>. [Accessed 31 May 2020].
- [2] <https://www.reuters.com/article/us-health-coronavirus-bangladesh-idUSKBN20V0FS>. [Accessed: 31-May- 2020].
- [3] "Coronavirus: Bangladesh declares public holiday from March 26 to April 4", Dhaka Tribune, 2020. <https://www.dhakatribune.com/bangladesh/2020/03/23/govt-offices-to-remain-closed-till-april-4>. [Accessed: 30- May- 2020]
- [4] "Bangladesh logs highest weekly Covid-19 deaths, cases", Dhaka Tribune, 2020. <https://www.dhakatribune.com/bangladesh/2020/05/30/covid-19-deaths-top-600-cases-top-44-000>. [Accessed: 30- May- 2020]
- [5] "IEDCR", [iedcr.gov.bd](http://www.iedcr.gov.bd), 2020. <https://www.iedcr.gov.bd/>. [Accessed: 30-May- 2020]
- [6] "Covid-19: Bangladesh likely to end general holiday on May 31", Dhaka Tribune, 2020. <https://www.dhakatribune.com/bangladesh/2020/05/27/state-minister-update-of-lockdown-coming-soon?>. [Accessed: 30- May- 2020].
- [7] Walaa Medhat, Ahmed Hassan, Hoda Korashy. "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, pp. 1093–1113, 2014.
- [8] D. Kang and Y. Park, "Measuring customer satisfaction of service based on an analysis of the user generated contents: Sentiment analysis and aggregating function based MCDM approach," 2012 IEEE International Conference on Management of Innovation & Technology (ICMIT), Sanur Bali, 2012, pp. 244-249, doi: 10.1109/ICMIT.2012.6225812.
- [9] Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, Hsuan-Shou Chu. "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," Knowledge-Based Systems, vol. 41, pp. 89–97, 2013.
- [10] Michael Hagenau, Michael Liebmann, Dirk Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," Decision Support Systems, vol. 55, pp. 685–697, 2013.
- [11] Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena. LargeScale Sentiment Analysis for News and Blogs. ICWSM'2007 Boulder, Colorado, USA.
- [12] LORI YOUNG, STUART SOROKA, "Affective News: The Automated Coding of Sentiment in Political Texts," Political Communication, vol. 29, no. 2, pp. 205-231, 2012.
- [13] Hamed Jelodar, YongliWang, Rita Orji, Hucheng Huang. Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. arXiv preprint arXiv:2004.11695, 2020 - arxiv.org.
- [14] Qanita Bani Baker, Farah Shatnawi, Saif Rawashdeh, Mohammad Al-Smadi, Yaser Jararweh."Detecting Epidemic Diseases Using Sentiment Analysis of Arabic Tweets." Journal of Universal Computer Science, vol. 26, no. 1, pp. 50-70, 2020.
- [15] "Arturo Montejo-Ráez, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López. Ranked WordNet graph for Sentiment Polarity Classification in Twitter." Computer Speech and Language, vol. 28, pp. 93–107, 2014.
- [16] Xuehua Han, JuanleWang, Min Zhang, XiaojieWang. Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. Int. J. Environ. Res. Public Health 2020, 17, 2788;doi:10.3390/ijerph17082788. www.mdpi.com/journal/ijerph.
- [17] Nikhil Kumar Rajput, Bhavya Ahuja Grover, Vipin Kumar Rathi. WORD FREQUENCY AND SENTIMENT ANALYSIS OF TWITTER MESSAGES DURING CORONAVIRUS PANDEMIC.arXiv:2004.03925v1 [cs.IR] 8 Apr 2020.
- [18] A. S. M. Kayes, Md. Saiful Islam, Paul A. Watters, Alex Ng, Humayun Kayesh. Automated Measurement of Attitudes Towards Social Distancing Using Social Media: A COVID-19 Case Study. doi:10.20944/preprints202004.0057.v1.
- [19] Long Chen, Tongyu Yang, Jiebo Luo, Hanjia Lyu, Yu Wang. In the Eyes of the Beholder: Sentiment and Topic Analyses on Social Media Use of Neutral and Controversial Terms for COVID 19. arXiv:2004.10225v1 [cs.SI] 21 Apr 2020.
- [20] Abdul Kader Mohiuddin. A Pandemic Review of Covid-19 Situation In Bangladesh.ISSN (e): 2689-1026, Website: <http://usajournalshub.com/index.php/TAJMSPR>.
- [21] Kazi Abdul Mannan and Khandaker Mursheda, Farhana. Knowledge and perception towards Novel Coronavirus (COVID 19) in Bangladesh. MPRA Paper No. 99656, posted 15 Apr 2020. Online at <https://mpra.ub.uni-muenchen.de/99656/>
- [22] Md. Shahidul Islam, Jannatun Irana Ira, K. M. Ariful Kabir, Md. Kamrujjaman. Effect of lockdown and isolation to suppress the COVID-19 in Bangladesh: an epidemic compartments model. doi:10.20944/preprints202004.0193.v2.
- [23] Noah C Peeri , Nistha Shrestha, Md Siddikur Rahman, Rafdzah Zaki, Zhengqi Tan, Saana Bibi, Mahdi Baghbanzadeh, Nasrin Aghamohammadi, Wenyi Zhang, Ubydul Haque. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? International Journal of Epidemiology, 2020, Vol. 0, No. 0. doi: 10.1093/ije/dyaa033.
- [24] Kim, Y., 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.
- [25] Johnson, Rie, and Tong Zhang. "Effective use of word order for text categorization with convolutional neural networks." arXiv preprint arXiv:1412.1058 (2014).