LSTM-RNN Based Sentiment Analysis to Monitor COVID-19 Opinions using Social Media Data

Ghaida Alorini and Danda B. Rawat
Data Science and Cybersecurity Center (DSC[2])
Dept. of Electrical Engineering & Computer Science
Howard University, Washington DC, USA
Ghaida.Alorini@bison.howard.edu, danda.rawat@howard.edu

Dema Alorini
Comm College, Computer Science & Info Technology
Princess Nourah bint Abdulrahman University
Riyadh, Saudi Arabia
dalorini1@gmail.com

ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500897

***Abstract*—Lately people are using social media for sharing their thoughts, insights about different topics and issues. The main aim of social media is to connect users and update their statuses/thoughts. One of the most used online social networking sites to share information is Twitter with roughly 330 million uses globally and 48.35 million US users where they share their opinions and thoughts. Recently, the world faced a serious pandemic, Corona virus disease (COVID-19) outbreak and the World Health Organization (WHO) declared the virus as a global health emergency. The COVID-19 started in late December of 2019 in Wuhan City, Hubei Province, China. Around the world during this time, individuals use social media to share their opinions about the pandemic. Because of the lack of information about the virus, people switched to micro-blogging platforms such as Twitter. In this study, we utilize natural language processing (NLP) techniques for opinion mining to extract negative and positive sentiments/tweets on COVID-19. We investigate NLP based sentiment analysis using Recurrent Neural Network (RNN) model with Long-Short Term Memory networks (LSTMs). Predicted sentiment using LSTM-RNN, which gives high accuracy, can be used to educate people about the virus.**

***Index Terms*—Natural Language Processing, LSTM, Recurrent Neural Network, Coronavirus, COVID-19, Sentiment.**

I. Introduction

Corona virus disease (COVID-19) epidemic has spread from Wuhan, China to almost 214 countries across the world. On 30th January 2020, the World Health Organization (WHO) declared the Chinese outbreak of COVID-19 a Public Health Emergency of International Concern particularly for countries with vulnerable population and health systems. According to the WHO, this new virus continues to spread through direct or indirect close contact with infected individuals. The spread of the virus can be interrupted by early detection, isolation, and prompt treatment [1, 19, 20]. Typically, when a disease outbreaks it grabs the public attention, misguided information and recommendations are often given to people without a reliable source. The sudden outbreak of the virus has released what the U.N. Secretary-General Antonio Guterres called a "pandemic of misinformation". In the early months of 2020, information and news of corona virus were published and shared on social media. Social networking has played a huge role in making the public aware of the virus outbreak. However, social media users are spreading misinformation, which resulted in panic and affected the mental health of other social media users. The field of infidomalgy has studied information patterns on social media for many years, and the COVID-19 pandemic has been referred as the first social media infodemic [2]. Social media infodemic has spread panic and affected many social networking users emotions.

In this paper, we leverage the Natural Language Processing (NLP) techniques for opinion mining to extract negative and positive tweets on COVID-19. We investigate NLP based sentiment analysis using the Recurrent Neural Network (RNN) model with Long-Short Term Memory networks (LSTMs). The workflow of this paper starts with data collection using Twitter Application Programming Interfaces (API). Following this step, data is pre-processed and tokenized. This is a crucial step since, without clean data, the dataset is a cluster of words that the computer cannot understand. For instance, tweets with stopwords that are irrelevant to help identify the text. After preparing data, we perform Exploratory Data Analysis (EDA) to analyze and model our dataset visually to extract important information. Finally, we apply NLP techniques for sentiment classification and deep learning based on RNN with LSTM learning model. *Implementing the LSTM algorithm to our data provides better results for predicting tweet sentiments. The output predictions of the model are compared to Textblob results to show sufficiency in deploying the methodology to the dataset.

II. Background and Literature Survey

Our goal is to levarage the deep learning for sentiment analysis based on COVID-19 tweets. Deep learning is a subset of machine learning which is based on artificial neural networks ([21]). It's a popular method for sentiment and semantic analysis to analyze text content. Recurrent neural networks are widely utilized in NLP and it has shown great efficiency in building language models. Many researchers use different deep learning methodologies on social networking sites such as Twitter, Reddit [3] and health websites [4]. For example work in [5], applied deep learning methods on social media data to design a software system, that detects the outbreak of influenza-like illness symptoms that can be confirmed by existing official sources. Tadesse et al [6], presented a study that applies both Long Short Term Memory (LSTM) neural network and Convolutional Neural Network (CNN)

978-1-7281-7122-7/21/$31.00 ©2021 IEEE

learning methods to detect suicidal posts through the social microblogging site Reddit. Hu et al [7] investigated Adverse Drug Reactions (ADR) from online health social networks using deep learning related methods such as Long Short Term Memory (LSTM) neural network and conditional random fields (CRFs) for ADR mention recognition. Du and colleagues [8] studied the comparison of machine learning and deep learning methods to extract psychiatric stressors from Twitter and indicated the use of deep learning as the best mechanism. Similarly, in previous studies, online networking platforms are useful to extract health related data. Our proposed study uses Twitter as a social media platform to extract opinions from COVID-19 related tweets. COVID-19 comments from media platforms are useful to extract information to better understand opinions, insights and emotions regarding the global pandemic. However, many people share different point of views about the virus, to help evaluate the comments regarding Corona virus, NLP techniques are utilized to classify emotions and opinions.

III. Framework and Methodology

A. Data Gathering

In the past decade, social media platforms that facilitate interactions between users have been rising and slowly becoming a necessity in people's lives. Numerous researchers have been using social media data to conduct machine learning experiments. Few of these social networking services are microblogging platforms, which is a new form of communication for users to share their current status in short posts. Twitter, one of the largest microblogging sites has seen rapid growth since it launched in October 2006. Twitter users are mostly sharing opinions, thoughts and insights on current news and issues happening around the world which is a great source to evaluate and analyze targeted topics using machine learning techniques. Twitter offers data through Application Programming Interfaces (APIs) or existing Twitter data sets. This feature provides different tools and libraries to enable third-party developers to glance at billions of tweets. The API service has been a great advantage to researchers, governmental institutions, and companies. The procedure requires an authentication process for third-party developers to gain access to tweets data. After completing the authentication process, access tokens are obtained for data authorization. For this NLP study, the best strategy is to work with data through Twitter API for automatic embedded content. An open source Python package called Tweepy is a convenient library to access Twitter contents through API's. The authentication process requires signing up for a developer account to generate API keys, consumer token, and access key. The keys allow us to interact with data by collecting tweets that contain keywords, hashtags, or specific users. API Python libraries help stream tweets based on desired topics between specific dates, for instance, the last two weeks of tweets containing corona virus. Fig. 1 shows the typical flow model.

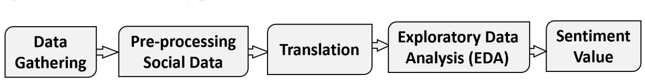

Fig. 1: Flow/System Model.

B. Preprocessing Social Data

Social networking sites contain unstructured data and irrelevant texts. The goal of preprocessing is to perform text mining to extract relevant information from unstructured data which helps transform the data into a clean format to employ for sentiment analysis. Data cleaning is done using natural language processing toolkit (NLTK) written in Python. that has interfaces, libraries and lexical resources such as Wordnet. During data preparation, data comes in two formats, a corpus and document-term matrix. The first format is the corpus which is a collection of texts put in a table. To obtain the corpus, we use Pandas, an open-source data analysis and manipulation tool built on top of Numpy. This tool is built on top of the Python programming language. Python Pandas offer an object called a dataframe that organizes information in a table. Every row in the dataframe has an ID and every column has the same datatype. Fig. 2 is an example of tweets containing the topic "Coronavirus" put in a dataframe. However, the corpus contains a lot of unnecessary data such as stopwords, URLs, punctuation, etc. Data must be cleaned to remove irrelevant texts from the corpus. Stop words can be filtered out before or after the processing of natural language text [9].

To obtain our data in matrix form, we clean and tokenize the text. The initial step to obtain clean tweets is to remove punctuation, symbols, stopwords, and URLs. Also, to make data more simple, we convert all letters to lower case (which does not change the semantics/meaning of the word) using Python regular expressions, available through 're' module. Processing takes raw input as tweets, removes noise and produces an output of clean data. Using the re module we can specify rules for possible strings to match. For instance, removing URLs from the middle of tweets using a 're.sub' function that removes URL links from the string. Also, to reduce noise, we focused on relevant hashtags that contain"Coronavirus" and "COVID19". Tokenization in NLP is the process of breaking text corpus into meaningful elements called tokens that could be either by word or sentence for lexical analysis [10]. There are two functions used to tokenize texts either by unigrams or bigrams, or both. Previous studies for different context found that using two features for similar classification showed great performance [11]. In this process, both n-grams and bi-grams are used as features to break sentences. Our approach requires an analysis of opinions and emotions in tweets, which indicates that the sentiment classification is important. Bi-gram is a function where context matters which is useful for sentiment classification and analysis. For instance, the words "bad" and "hate" are counted as negative in a sentence, however, if they are accompanied with "not" they are considered positive. On the other hand, N-grams takes each word by itself from the corpus and analyzes it. Unigrams are used to capture the order of words. After organizing the data into bigrams and unigrams it's easy to see how often certain words are preceded. The natural language processing toolkit includes Textblob a Python library that offers a simple API to access its credentials. The

	tweet_dt	topic	id	username	name	tweet	like_count	reply_coun
0	2020-10-16	Coronavirus	1316948621674356736	CMcKNichols	Christopher Nichols	No only has the U.S. botched the response in comparison to other nations, they know it. The world consensus is that… https://t.co/Cldv2bxXiT	0	
1	2020-10-16	Coronavirus	1316948606029615106	Y_Alkhalidi	يوسف الخالدي - Yousef AlKhaldi	US House Speaker @SpeakerPelosi writes letter to Democratic Representatives on #COVID_19 #coronavirus #stimulus pro… https://t.co/cV2BsOdJc7	0	
2	2020-10-16	Coronavirus	1316948584386957312	LastGreatAct	LastAct	@NRafter @meganranney This is more of what I see happening.\n\nhttps://t.co/wQzv3NTMAK	0	
3	2020-10-16	Coronavirus	1316948574211538944	WitchieWitch404	✵Witchie Witch✵	We're trapped here😿 \nFox News: Trudeau: US-Canada border to remain closed until coronavirus is under control.… https://t.co/zaxzzQdn1F	0	
4	2020-10-16	Coronavirus	1316948573381140482	KathrynLEngel	Kathryn L Engel	@JoeBiden And, weeping, you MUST protect all of us https://t.co/PWC6Y3cAzR	0	

Fig. 2: Example Twitter dataframe.

```
Filtered Sentence: ['Coronavirus', 'hurting', 'store', 'traffic', 'e-commerce', 'wo', "n't", 'help', 'much']
Stemmed Sentence: ['coronaviru', 'hurt', 'store', 'traffic', 'e-commerc', 'wo', "n't", 'help', 'much']
```

Fig. 3: Stemming tokenization.

library tokenizes text by splitting standard contractions using the 'nltk.tokenize.mwe' module. This module merges word expression into single tokens to create a list of words from a string. Another feature that NLTK offers is called stemming which is a process of reducing a word to its root word. Fig. 3 shows an example of tweet tokenization after applying the stemming process of linguistic normalization. The output after cleaning our data generates a document term matrix to later be used for machine learning processing and analysis.

C. Translation

Coronavirus disease (COVID-19) has impacted almost 215 countries across the world. During the pandemic, many people across the globe use social networking platforms to express emotions and share opinions about the virus. We translated tweets in other language such as Arabic language to English by using Yandex Translator API that translates more than 90 languages including Arabic. This translator requires an API key, which can be obtained by registering with the Yandex website. Following obtaining API keys, installation is done for Python programming language. [1]

D. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the process of investigating a data set to summarize its characteristics, which is mostly done visually. The methodology is mostly used to discover patterns, test assumptions, and spot abnormalities with statistics and graphic representation. The quantitative data analytic tradition is based on the original work of John Tukey [12]. The main goal of EDA is to develop a clear understanding of our data. The EDA input includes both the corpus and document-term matrix. The process consists of three simple steps. The first step is to generate questions about your data. For instance, a question to consider about the data is whether there are any missing values and what type of variations occurs. Another example question is what are the word frequencies and word counts for tweets. After setting the goals for investigating data, we can start the visualization and modeling process using different Python open-source libraries. The second step of EDA is to search for answers by visualizing and modeling data. The first concern regarding obtained data is whether there are any missing values found. Missing values can lead to inaccurate inferences and conclusions. To solve this issue, we use Box-plot through Seaborn which is a Python library for making statistical visualizations. Boxplot uses IQR method to visually display data along with its outliers. The Seaborn Python library makes it easy to identify outliers data points that are unusual. This visualization technique gives an indication of how values in the data are spread out. Fig. 4 is an example visualization of missing values using Boxplot. To solve the issue of missing values, outliers are removed and restored to make it less prominent.

For this study, coronavirus tweets are the main target to observe and analyze. Before conducting a sentiment analysis, the distribution of variables can be beneficial to visualize and observe. Variables such as the quantity of words used by a certain group of users can be performed and observed visually. For instance, observing the top relevant words used in covid related tweets can help define the vocabulary of preferred words in the Bag of Words (BoW) model, as shown in Fig. 5.

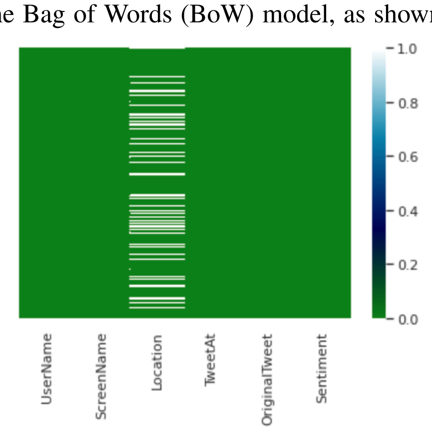

Fig. 4: Statistical visualizations of missing values

[1]https://yandex.com/dev/translate/

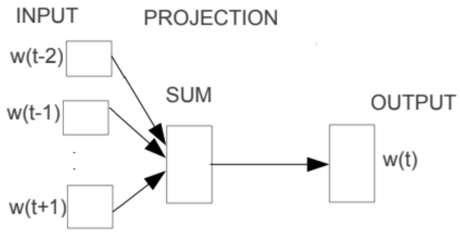

Fig. 5: Bag of Words (BOW) architecture.

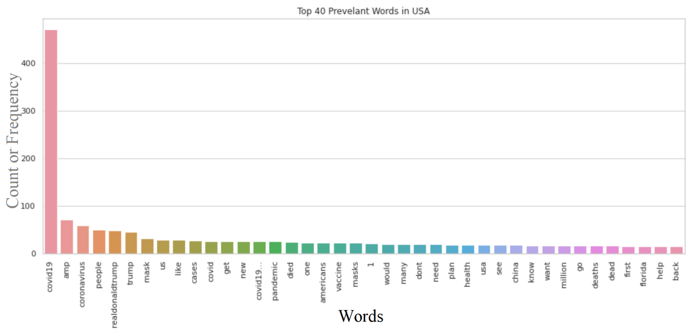

Fig. 6: Relevant words for coronavirus tweets.

Typical top relevant words for coronavirus tweets in the USA are shown in Fig. 6. We observe that the most frequent words used in tweets are vaccine, pandemic, realDonaldtrump and deaths.

IV. Sentiment Analysis

Sentiment analysis (or opinion mining) process uses NLP and machine learning to find the polarity of text or sentiments such as positive, neutral, and negative. Sentiment analysis has been handled at the phrase level most recently [13][14]. Before, it was mostly handled on a sentence or document level. The two types of techniques used for sentiment analysis are lexicon-based techniques and machine learning based techniques [15]. Machine learning technique features consist of Parts of Speech (POS) tagging, n-grams, bigrams, and Bag of words modeling. The analysis process is performed using the corpus as an input. The document term matrix is not used to find sentiments since order matters in a sentence. To find sentiments, we import Textblob an open-source library in Python built on top of NLTK, that supports complex analysis of data. The Textblob library utilizes NLTK (Wordnet) interface, a lexical database of English words linked using semantic relationships. Output of the process is an overall polarity score and subjectivity score. The polarity score ranges from '-1' for negative and '+1' for positive sentiments, while subjectivity score ranges from 0 for objective and 1 for subjective (opinion) sentiments. We calculated the overall sentiment of tweets content as negative, neutral or positive. Tweets with a score of 1 are labeled as positive and tweets with a score of -1 are labeled negative.

A. Word Embedding

Word embedding is the process of converting texts to numbers. The word embedding tool utilizes both the BOW model and skip-gram model for computing vector representations of words [17]. Word2Vec is a popular natural language processing method that uses neural networks to process texts by vectorizing it to numbers. Word optimization, learning and data accuracy can be achieved using this tool. One of the two main algorithms in Word2Vec is the BOW model, which has better accuracy for frequent words than the Skip-gram model. In a previous study, word vectors with semantic relationships were found to improve the NLP process of information retrieval and machine translation [18]. By implementing the BOW methodology distribution is used by combining the surrounded words to predict the word in the middle. The process input is the corpus and the output is a set of vectors (numbers). Results can be used as a feature for natural language processing using deep learning neural networks. Fig. 5 is an example of bag of words (BOW) architecture of combining surrounding words to predict the word in the middle [18].

V. Long Short Term Memory (LSTM) Recurrent Neural Network

Deep learning processes consist of several effective and popular models that are used to solve a variety of problems [16]. Neural networks computing is performed using sentiment analysis. For this study, we leveraged the deep learning architecture based on the LSTM recurrent neural network, a type of recurrent neural network that has previously shown great success in language-related tasks in different context.

Before applying the LSTM network model, data is preprocessed and labeled with sentiment values. The main target for this process is to map the sentiment label and tweet input. The sentiment label includes extremely positive, extremely negative, negative, neutral, and positive sentiments. The training set for learning model had 18,230 COVID-19 related tweets which is 80% of data, while the test dataset for the model had 4,557 tweets with 20% of data. We constructed the LSTM-RNN model for this training and testing. Next, we added three layers to the network model including the embedding layer, LSTM layer, and dense layer. Then, we added a dropout layer to help prevent overfitting of 20% of neurons. The LSTM layers use the input to make predictions to produce an output of predicted values that is close to the actual values.

Our approach based on the LSTM-RNN model achieved an accuracy score of 86% , which is considered a high score. The contribution of semantic and sentiment analysis techniques in deep learning achieved the desired output for positive and negative tweets and a high accuracy score. To compare the results of LSTM neural network, we deploy the model on two datasets. The first dataset contains COVID-19 confirmed cases in the US and the second dataset contains coronavirus tweets for 14 days. For covid cases, the LSTM model observes the last 7 days to forecast 14 days ahead of cases. This process is achieved using the Tensorflow Keras library. Before using the model, data is split into small sequences for training and test data. Preparation is done for timesteps, features and samples. After preparing sequences, the model is trained and evaluated. Fig. 7 shows the results of COVID-19 cases prediction based on LSTM model. *Processing methodology depends on the past 250 days of covid cases. The prediction results shows an increase of cases in the US for the 14 days ahead.

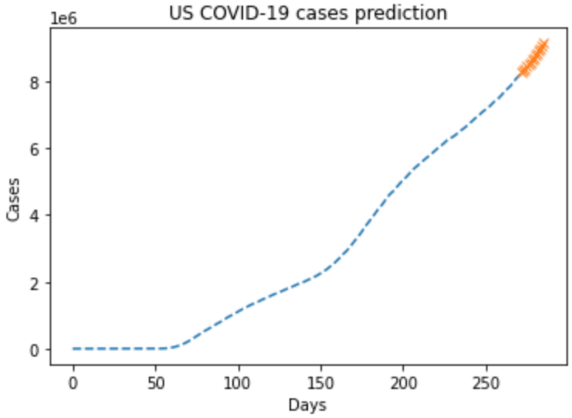

Fig. 7: COVID-19 new cases prediction for next 14 days

After observing new covid cases, we used LSTM-RNN model for tweet based sentiment analysis during the same period and used the confusion matrix for evaluation. A comparison of actual sentiments and predicted sentiments are shown in Fig. 8 where each row represents instances of the predicted class and true class. To visualize the algorithm performance, 'Sklearn.metrics' is imported for evaluation using the Scikit-learn framework. The results have shown a clear indication of how sentiment predictions are close to the true sentiment, as shown in Fig. 8. Next, we evaluate sentiments for specific countries depending on the location of the tweet on the dataset. The process requires using Pandas for indexing and selecting data. For instance, selecting the "Location" column to select specific countries. For indexing, columns are not filtered, however, rows are retrieved depending on the location of the user. After obtaining the right information, a confusion matrix is deployed to evaluate actual sentiments and sentiment predictions for a specific country. Fig. 8 (upper part) is an example visualization for sentiment prediction using our complete dataset for all countries and Fig. 8 (lower part) is an evaluation of US only. We compared different countries (USA, UK, Canada - all plots are not included in this paper), extremely negative sentiments are mostly found in the USA, and extremely positive sentiments are found in Canada.

To further compare the LSTM/RNN model predictions and the original sentiment analysis using Textblob, we implement two pie charts for visualization. The main objective is to display sentiment values extremely positive, positive, neutral, negative, extremely negative. Sentiment values for both methodologies are compared using pie charts, which is generated using a Python library called matplotlib to plot graphs. The dataset obtained using Textblob contains the original sentiments for COVID-19 related tweets using the polarity and subjectivity score. However, the updated dataset from LSTM prediction contains predicted sentiments and actual sentiments from the sentiment analysis. To examine the accuracy of the deep learning model prediction, pie charts are utilized to observe and visualize. Fig. 9 shows an example

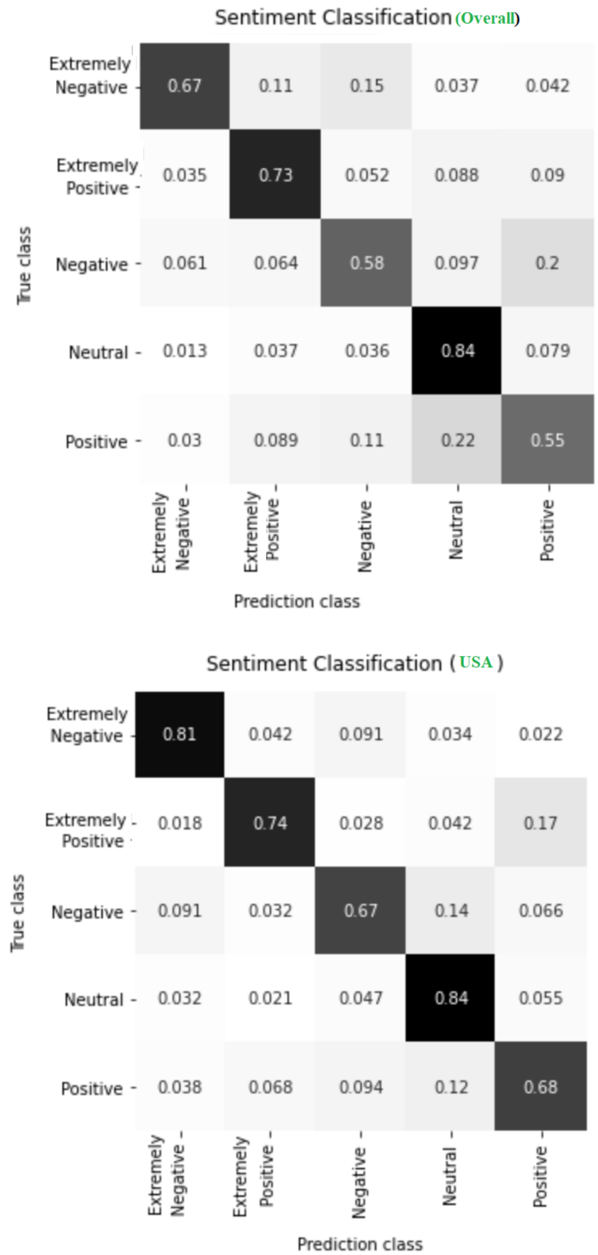

Fig. 8: LSTM RNN based sentiment predictions (overall and USA).

of actual sentiments compared wth the LSTM-RNN based predictions using pie charts. We observe that actual sentiments and new sentiment predictions are close in terms of scores, as shown in Fig. 9. The negative score for COVID-19 tweets is the highest from both LSTM prediction and actual sentiments analysis. These results of deploying the evaluation matrix and comparison of sentiments values showed how well is the overall performance of the LSTM-RNN model for analyzing sentiment related to COVID-19.

VI. Conclusion

This paper has analyzed and evaluated COVID-19 related opinions using social media (e.g., Twitter) data. The main goal of this paper was to detect whether COVID-19 related

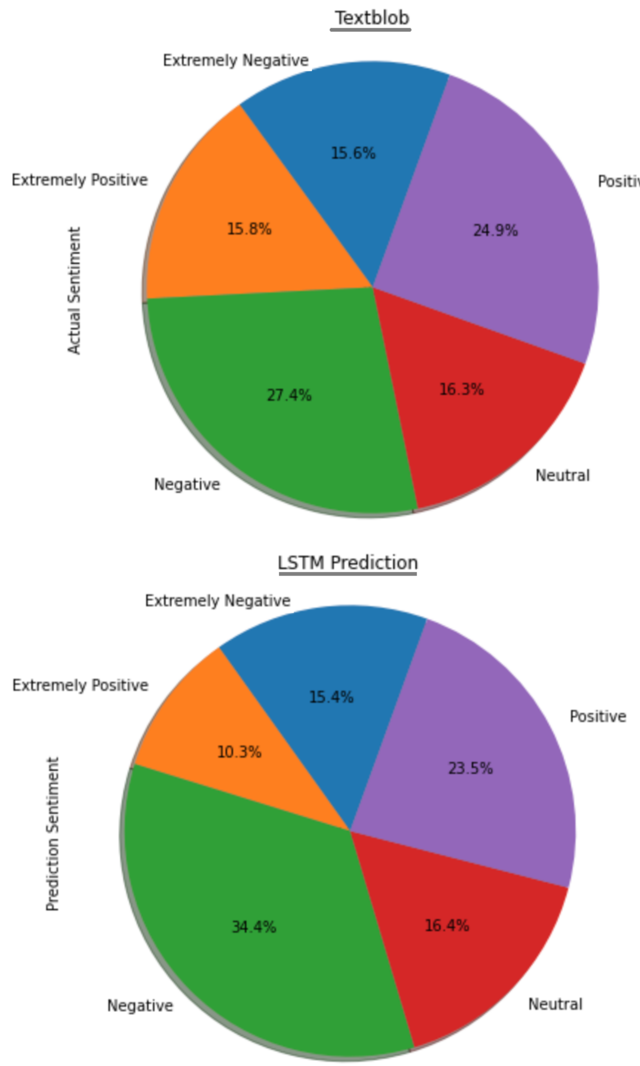

Fig. 9: Comparative Analysis of TextBlob and LSTM for COVID-19 related sentiments.

tweets are positive or negative. NLP assisted LSTM-RNN based sentiment analysis techniques are deployed to achieve higher accuracy for prediction. To find the sentiment polarity, word embedding model (Word2Vec) was used to process a desired output of vectors for social media data. The proposed approach resulted in better prediction for sentiments (extremely negative, extremely positive, etc.) that was close to the ground truth. This estimated sentiment can be used for counseling or educating people so that they can be better prepared for protecting them and others. Our future works include sentiment analysis related to COVID-19 vaccines.

Acknowledgments

This work was supported in part by the US NSF under grants IIS-1924092 and 1828811, and by the DoD Center of Excellence in AI/Machine Learning (CoE-AIML) at Howard University under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory. Ghaida Alorini's work was supported by SACM, Washington DC, USA.

References

[1] World Health OrganizationNovel Coronavirus(2019-nCoV), Situation Report – 12 (2020)

[2] Ahmad, A.R.; Murad, H.R. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: Online questionnaire study. J. Med. Internet Res. 2020, 22, e19556.

[3] Jelodar, Hamed & Wang, Yongli & Orji, Rita. (2020). Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. 10.1101/2020.04.22.054973.

[4] Xia L., Wang G.A., Fan W. (2017) A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media. In: Chen H., Zeng D., Karahanna E., Bardhan I. (eds) Smart Health. ICSH 2017. Lecture Notes in Computer Science, vol 10347. Springer, Cham.

[5] Serban, O. et al. "Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification." Inf. Process. Manag. 56 (2019): 1166-1184.

[6] Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L. Detection of Suicide Ideation in Social Media Forums Using Deep Learning. Algorithms 2020, 13, 7.

[7] Buzhou Tang, Jianglu Hu, Xiaolong Wang, Qingcai Chen, "Recognizing Continuous and Discontinuous Adverse Drug Reaction Mentions from Social Media Using LSTM-CRF", Wireless Comm. and Mobile Computing, vol. 2018, Article ID 2379208, 8 pages, 2018.

[8] Du, Jingcheng et al. "Extracting psychiatric stressors for suicide from social media using deep learning." BMC medical informatics and decision making vol. 18,Suppl 2 43. 23 Jul. 2018.

[9] ShaidahJusoh , Hejab M. Alfawareh, Techniques, Applications and Challenging Issues in Text Mining, Intl J. of Computer Science Issues, Vol. 9, Issue 6, No 2, November -2012 ISSN: 1694-0814

[10] C.Ramasubramanian , R.Ramya, Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2013

[11] Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., and Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. Proc. of ISCRAM, Baden-Baden, Germany.

[12] Behrens, J.T., DiCerbo, K.E., Yel, N. and Levy, R. (2012). Exploratory Data Analysis. In Handbook of Psychology, Second Edition (eds I. Weiner, J.A. Schinka and W.F. Velicer). doi:10.1002/9781118133880.hop202002

[13] Wilson, J. Wiebe, and P. Hoffman. 2005. Recognizing contextual polarity in phrase level sentiment analysis. ACL.

[14] Apoorv Agarwal, Fadi Biadsy, and Kathleen Mckeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 24–32, March

[15] J. Singh, G. Singh, and R. Singh, A review of sentiment analysis techniques for opinionated web text, CSI Trans. ICT, 2016.

[16] X. Ouyang et al, Sentiment Analysis Using Convolutional Neural Network, 2015 IEEE Int. Conf. on Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput, pp. 23592364, 2015.

[17] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS '13), pp. 3111–3119, December 2013.

[18] Mikolov, Tomas & Corrado, G.s & Chen, Kai & Dean, Jeffrey. (2013). Efficient Estimation of Word Representations in Vector Space. 1-12.

[19] Pranav Singh, Sukumar Nandi, Kayhan Ghafoor, Uttam Ghosh and Danda B. Rawat "Preventing COVID-19 Spread using Information and Communication Technology." IEEE Consumer Electronics Magazine (2020) DoI: https://doi.org/10.1109/MCE.2020.3047608

[20] Halgurd Maghdid, Kayhan Zrar Ghafoor, Ali Safaa Sadiq, Kevin Curran, Danda Rawat and Khaled Rabie "A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: Design study." 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI). 2020.

[21] F. Olowononi, D. B. Rawat, and C. Liu. "Resilient Machine Learning for Networked Cyber Physical Systems: A Survey for Machine Learning Security to Securing Machine Learning for CPS." IEEE Comm Surveys & Tutorials (2020). DoI: https://doi.org/10.1109/COMST.2020.3036778