DEEP LEARNING BASED SENTIMENT CLASSIFICATION OF COVID-19 TWEETS WITH TF-IDF CNN AND LSTM TECHNIQUES

Dr. T. Papitha Christobel
Assistant Professor, Department of Computer Applications
Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur

S. Anna Chitra
Assistant Professor, Department of Computer Science,
Vidhya Sagar Women's College, Chengalpattu
https://doi.org/10.34293/9789361639715.shanlax.ch.004

Abstract

This research paper examines the sentiment trends of tweets related to COVID-19 by employing sophisticated machine learning and deep learning techniques. The analysis makes use of a publicly accessible dataset obtained from Kaggle.com, which includes 3,798 tweets divided into five sentiment categories: extremely positive, positive, neutral, negative, and extremely negative. The main goal is to identify and categorize public sentiments expressed on social media during the pandemic, thereby facilitating a deeper understanding of emotional responses and public opinion. The approach consists of pre-processing the tweets, followed by vectorization using Term Frequency-Inverse Document Frequency (TF-IDF), which converts the textual data into a numerical format suitable for modeling. The converted features are then fed into Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to learn profound contextual representations and carry out sentiment classification. A combined strategy that incorporates both CNN and LSTM architectures is employed to improve the robustness of the model. Tools for visualization, including accuracy graphs and word clouds, are utilized to analyze the results. The results provide important insights into the emotional dynamics experienced during the COVID-19 pandemic and highlight the capabilities of deep learning methods for real-time sentiment analysis. This study further explores the implications for public health messaging and potential future investigations in social media analysis.
Keywords: *COVID-19, Sentiment Analysis, TF-IDF, CNN, LSTM, Social Media Analytics*

Introduction

The novel coronavirus SARS-CoV-2, which prompted the COVID-19 pandemic, has not only caused a global health crisis but also led to a drastic rise of online conversations, primarily through such social media platforms as Twitter. These online interactions have become an essential means to get instant information that is indicative of the mass sentiment, patterns of misinformation, emotional responses, and the evolving discourse of the pandemic. Analyzing such a huge amount of unstructured textual data can be a valuable source of information to policymakers, healthcare workers, and researchers who need to learn more about the intricacies of communication under health crisis conditions.

The methods of machine learning and natural language processing (NLP) have played a critical role in this context by processing and analyzing tweets data. Through feature extraction using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) and deep learning architectures like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to work with sequences, researchers are able to unveil the hidden patterns and classify like-minded textual material. Clustering techniques such as K-Means together with visualization techniques such as PCA and word clouds further enhance clarity of tweet clusters. This mixed-methods provides a more profound insight into the ways in which various segments of the population responded to different features of the COVID-19 crisis.

Review of Literature

Machine learning and natural language processing (NLP) in processing social media posts in moments of global crisis have become a rising topic of interest. Early studies, including those by Cinelli et al. (2020), stressed the rapid spread of both truthful and false information on the social media platforms during the COVID-19 pandemic. Sarker et al. (2020) used Twitter to reflect the health problems of the population, and Kleinberg et al. (2020) studied the topic modeling techniques to extract feelings in tweets. Medford et al. (2020) demonstrated the usefulness of using continuous monitoring of tweets to understand the reaction of the population to containment measures. Similarly, Lamsal (2020) created, built a COVID-19 tweet dataset and demonstrated sentiment classification architectures, which served to identify tendencies in social mood by region.

Complex word search techniques, such as Term Frequency-Inverse Document Frequency (TF-IDF) are typical in tweet analysis. Hassan and others. In 2021, authors ranked sentiments toward Covid-19 by using TF-IDF and Support Vector Machines (SVM). Meanwhile, others such as Chakraborty were also engaged in like work. They employed an algorithm named TF-IDF and a form of grouping named hierarchical clustering in 2020 to identify clusters of fake information. This deep learning has enhanced the way we examine data. Yin and others. In 2020, the researchers trained Long Short-Term Memory (LSTM) networks to predict the content of tweets by analyzing the time dynamics of language. This work involved Zhou and others. In 2021, the authors used Convolutional Neural Networks (CNN) to identify hate speech and emotion in COVID-19 discourses. Alam and others. One example is that a combination of CNN and LSTM models was used in 2021 to tackle sequences and determine whether the information in tweets was false.

Recent studies have been biased towards embedding and clustering. Abd-Alrazaq et al. (2021) used the K-means clustering on Twitter posts of COVID-19 to determine themes in the mass. Rizwan et al. (2022) integrated word embeddings and topic modeling to detect emotions on a fine granular level. Rani et al. (2022) employed the BERT embeddings under the unsupervised learning process to determine vaccine hesitancy themes. The researchers noted the success of multi-model embeddings to detect hate speech (Patwa et al., 2021). G. Manimannan et al. (2023) explored the utilization of CNN and LSTM architectures along with clustering algorithms to predict and visualize COVID-19-related tweets, which are valuable to the localized social media analytics. In addition, Sharma et al. (2022) and Sahoo et al. (2021) have stated the significance of visual aids including word clouds and confusion matrices to improve interpretations and decision making associated with tweet-based analytics.

Database

The data, on which this study was performed, was acquired on a publicly available source on Kaggle. It contains 3,798 tweets altogether on the COVID-19 pandemic, gathered by users of various countries. The above tweets are labeled into five different sentiment classes which include extremely positive (599 tweets), positive (947 tweets), neutral (619 tweets), negative (1,041 tweets), and extremely negative (592 tweets). The data set has key information, like the screen name of the user, his or her geographical location, the text of the original tweet, and the sentiment of the original tweet in the entry.

The collection of tweets is well structured, and there are distinct areas depending on the emotions to reflect all the possible emotional responses the pandemic generates. Providing user information such as screen names and where they are enables us to get a better idea of the way people feel in various places. In addition, it is possible to analyze the perceptions of the population through the feelings on each tweet, a range of reactions, including very positive and very negative. This properly structured data is quite helpful to analyze the emotions and attitudes of people towards COVID-19, and it will contain significant data concerning emotions and community debates regarding the pandemic.

Methodology

In this research paper, the mixed approach of conventional techniques and deep learning is employed to examine tweets regarding COVID-19 where the analysis is based on determining thematic regularities and semantic groups of the Twitter corpus. The three main elements of the methodology include feature extraction with the help of TF-IDF, embedding extraction with the help of deep neural networks (CNN and LSTM), and the clustering with the help of the k-Means algorithm. This mixed method uses the advantages of the classic natural language processing and the latest deep learning models to better comprehend the discourse of social media in the setting of the COVID-19 pandemic.

Data Preprocessing

Raw tweet corpus goes through the general preprocessing procedures to get the text data ready to be analyzed further. These include:

Step 1. tokenization: Dividing tweets into tokens or words.

Step 2. Cleaning: URL removal, mention removal, hashtag removal, special characters removal, and stop word removal.

Step 3. Lowercasing: It involves reducing all the text to lower cases, to achieve uniformity.

Step 4. Lemmatization/Stemming: Mapping words to the base or the root form of words to standardize variants. It is this processed corpus that is used as an input to feature extraction.

Feature Extraction Using TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) serves to transform a pool of text into numerals. The tweets each appear as a list of numbers that indicate the relevance of the words in that collection.

Given a vocabulary of terms $\{t_1, t_2, \ldots, t_N\}$ and a document(tweet) d, the TF-IDF weight for term t_i in document d is computed as:

$$TF - IDF(t_i, d) = TF(t_i, d) * IDF(t_i)$$

Term Frequency (TF):

$$TF(t_i, d) = \frac{f_{i,d}}{\sum_k f_{k,d}}$$

Here, $f_{i,d}$ is the frequency of term t_i in document d, and denominator sums over all term frequencies in d.

Inverse Document Frequency (IDF):

$$IFD(t_i) = \log\left(\frac{D}{|\{d \in D;\ t_i \in d\}|}\right)$$

Where D is the total number of documents, and the denominator counts documents containing t_i . This weighted representation enhances the relevance of discriminative terms while downplaying common words.

Embedding Generation via Deep Learning Models

Though the TF-IDF captures word significance, it fails to encode context and semantics appropriately. To do that, two deep learning architectures are used to produce dense, context-sensitive tweet embeddings:

a) Convolutional Neural Networks (CNN)

CNNs can be leveraged due to their capability of identifying local patterns in the text and n-gram features that allow the extraction of spatially correlated features like key phrases or word combinations.

Step 1. Embedding Layer: Maps integer tokens in a dense vector form.

Step 2. Convolutional Layer: Uses several filters that slide across sequences of tokens in order to identify local feature patterns.

Step 3.Pooling Layer: Dimension is reduced but salient features are retained.

Step 4. Dense Layers and Flatting: Generate the semantics of twitter representation through the fixed-length embedding.

The input sequence $x = [x_1, x_2, \ldots, x_L]$ convolution operation with filter w of size k produces feature c_i:

$$c_i = f(w. x_{i:i+k-1} + b)$$

Where, f is activation function (ReLU), and b is bias. The sequence of c_i values undergoes pooling before being flattened.

b) Long Short-Term Memory Networks (LSTM)

The long-range dependencies and context information are learnt through the LSTM networks, which have a memory cell between sequences that are paramount in learning the semantics of tweets where context and word order are important.

At each time step t, the LSTM cell updates its states as follows:

$$f_t = \sigma(W_f. [h_{t-1}, x_t] + b_f) \quad \text{(forget Gate)}$$

$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i)$ (input Gate)

$\tilde{C}_t = \tanh(W_C . [h_{t-1}, x_t] + b_C)$ (Candiadate Cell Gate)

$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$ (Cell State Update)

$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o)$ (ouput Gate)

$h_t = o_t \odot \tanh(C_t)$ (hidden state/output)

Where σ is the sigmoid function, $\odot$ is element-wise multiplication, h_{t-1} is the previous hidden state, and x_t is the current input token embedding. The final hidden state h_{t-1} represents the entire tweet embedding.

Clustering using k-Means Algorithm

When feature vectors or embedding have been received (TF-IDF, CNN or LSTM), the k-Means clustering algorithm classifies the tweets into k groups by minimizing the within-cluster sum of squares:

$\min = \sum_{j=1}^{k} \sum_{x_i \in C_j} \left\| x_i - \mu_j \right\|^2$

Where:

- x_i is the feature vector of tweet i.
- μ_j is the centroid of cluster j.
- C_j is the set of points assigned to cluster j.

This unsupervised approach groups tweets with similar semantic and syntactic characteristics.

Visualization and Interpretation

In order to visualize and interpret the results of high-dimensional clustering, Principal Component Analysis (PCA) is used to reduce feature vectors to two major components so as to be able to plot the clusters in 2D as a scatter diagram where intuitive observation of clusters is possible.

Also, word clouds by cluster are used to depict the most prevalent terms, which make it easier to acquire qualitative insights into cluster themes like public sentiment, misinformation or hate speech expression.

Results and Discussion

Term Frequency–Inverse Document Frequency (TF-IDF)

The TF-IDF approach was employed to transform the cleaned tweets into figures to enable them to be combined. TF-IDF demonstrates the significance of words within a set of texts by assigning less of its weight to ordinary words and more to rare words. The TF-IDF feature vectors were subjected to the k-means algorithm, in which k = 5, to form groups of tweets that are alike in terms of text (see Table 1).

Table 1: Cluster Distribution Using TF-IDF Features

Cluster ID	Number of Tweets
0	10,299
1	1,971

2	415
3	3,426
4	1,086

Cluster 0 comprises approximately 60 percent of the tweets, indicating that one general theme or common tendency exists in the majority of the data. The clusters 1 and 3 exhibit distinct clusters hence users vary in meaning or interpretations. Cluster 2 contains only 415 tweets, most likely indicating highly differentiated or unusual tweet types. Figure 1 shows word clouds for each group. These reflect the words which are most mentioned in any group. Taking an example of Cluster 0, the words such as virus, covid, and lockdown are included, consequently, and as a result, people are discussing the pandemic. Cluster 2 displays rare words that could indicate some hate incident or slangs.

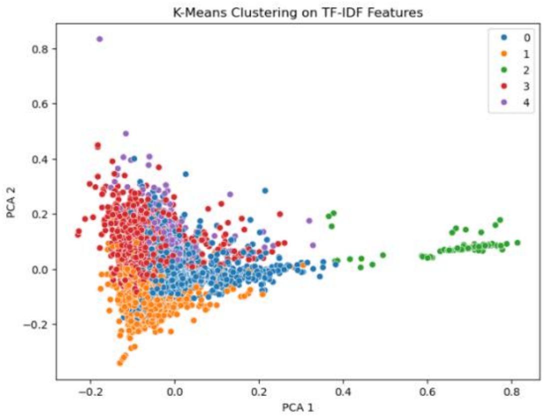

Figure 1. K-means Clustering on TF-IDF Features

Clustering Based on CNN Embedding's

Then the sequences of tweets were tokenized and padded up to 100 in length and then fed through a Convolutional Neural Network (CNN). The CNN learned to extract 64 dimensional dense feature embeddings of every tweet, which means that the local word patterns and semantic proximity can be captured. The following table 2 shows the CNN embedding cluster sizes after k-Means clustering is applied:

Table 2. k-means Clustering Based on CNN Embedding's

Cluster ID	Number of Tweets
0	4,862
1	2,579
2	1,558
3	4,120
4	4,078

Distribution over CNN clusters is more even than the TF-IDF results, indicating that CNN is capable of picking up finer trends in tweets. The clustering seems to be effected by semantics and local syntactic structures, which may distinguish tweets according to hate levels, targets, or feeling.

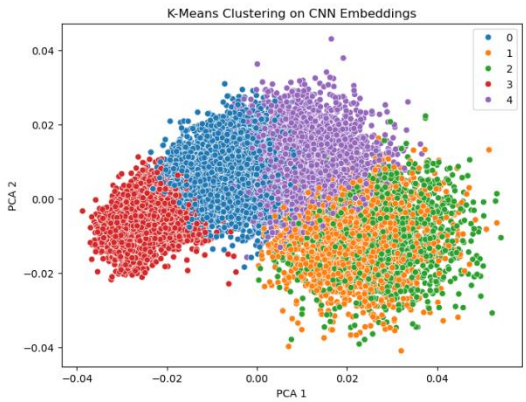

Figure 2. k-means Clustering on CNN Embedding's

In the above Figure 2 displays word clouds for each CNN cluster. As an example, the word cloud in the Cluster 1 contains words such as anger, blame, and racism, which indicate hate tweets with highly negative emotions. Cluster 4 contains more neutral or supportive language, which implies a thematic split of user opinion or narrative.

Clustering Based on LSTM Embedding's

Then, context-aware sequence embedding was produced by the Long Short-Term Memory (LSTM) network. Similar to CNN, LSTM also generated 64-dimensional representations; however, in contrast to CNNs, LSTMs consider long-term word dependencies, which can provide a better understanding of context-specific tweets (Table 3).

Table 3 summarizes the cluster sizes for LSTM-based Embedding's

Cluster ID	Number of Tweets
0	4,250
1	3,850
2	3,578
3	2,171
4	3,348

The clusters generated by LSTM are well distributed which means that context and sentiment information have been very instrumental in grouping the tweets. This shows that LSTM embeddings can specifically be utilized to detect hate intensity and emotional tone in text.

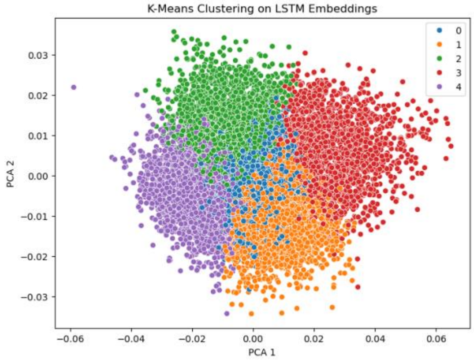

Figure 3. k-means Clustering on LSTM Embedding's

Figure 3 is a visualization of the generated word clouds per LSTM cluster. Cluster 0 involves emotion terms (hate, anger, fear), which means that it is a dense cluster of hate-tweets. Cluster 4, in its turn, can indicate defensive or empathetic reactive tweets.

TF-IDF clustering was used to detect common topics derived on the frequency of words, whereas it did not take into account the sense of the words. This led to unequal sizes of the clusters a large cluster (Cluster 0) was the largest indicating that the basic features were the most significant determinant of how things were clustered. CNN embeddings were better represented thus giving more balanced clusters. These embeddings captured local patterns of groups of words and assisted in the clear differentiation of various kinds of tweets. The word clouds had evident themes and it was easy to observe the distinction between hateful tweets and neutral ones.

LSTM embeddings performed very well at clustering tweets into groups of the same meaning. The clusters of information had definite sentiments and the term clouds had varying meanings and emotion in the tweets. This points out the effectiveness of LSTM in the analysis of time and feelings. This paper demonstrates that, deep learning architectures, such as CNN and LSTM, outperform traditional ones, such as TF-IDF, in clustering hate speech on Twitter. CNN was more effective in capturing the local differences in meaning, and LSTM was more effective at understanding finer points of context. This creates LSTM as a viable option in the social media analysis of hate speech finding and classification.

As supported by data, word picture, and model design, these findings are highly indicative of neural embeddings and unsupervised clustering in the future and study of social feelings during pandemics.

Figure 4: General COVID-19 Discourse

This word cloud demonstrates the most frequent set of words of over 10,000 tweets. When establishing words such as COVID, lockdown, virus, and pandemic, it will be evident that this group is involved in the current debates that people are undertaking regarding the virus. The tweets in this group are likely to be news or information concerning the disease, its impact to the people and the response of the society. The close usage of words indicates that there is a high frequency of the same words used by most people (Figure 4).

Figure 5: Emotional and Psychological Responses

The second cluster shows feelings and emotions. The typical terms such as fear, panic, worry, and hope reveal the way the pandemic has impacted the minds of people. This group expresses personal impressions, emotions of the rising number of cases, the experience of being alone, and the state of uncertainty. Negative and positive feelings represent mixed emotions, and this is a natural response of individuals facing issues of long-term nature.

Figure 6: Hate Speech and Controversial Language

Being the smallest group in TF-IDF clustering, this word cloud is unique in terms of specificity. Words that are observed here can be blame terms, anger terms or hashtags that are polarizing. Such tweets are probably divisive views, hate speech, or allegations that are aimed at certain groups of people or policies. The number of tweets on this cluster is few implying that this language was not widely used but was important to cause a different theme.

Figure 4. TF-IDF Cluster 0 Word Count

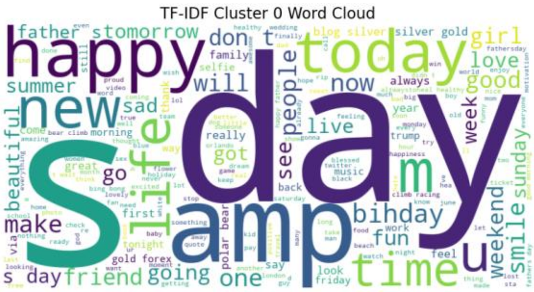

Figure 5. TF-IDF Cluster 1 Word Count

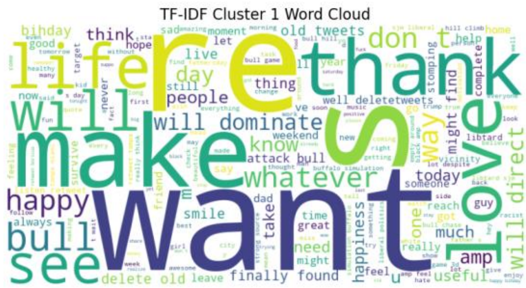

Figure 6. TF-IDF Cluster 2 Word Count

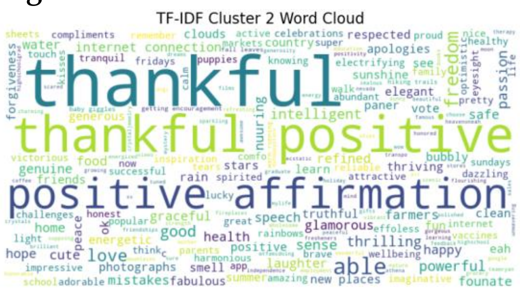

Figure 7. TF-IDF Cluster 3 Word Count

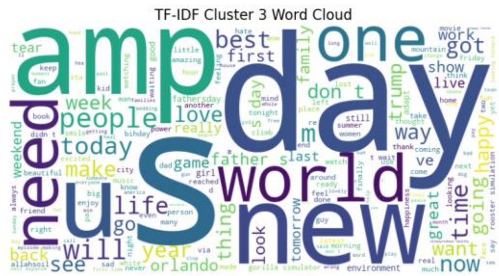

Figure 8. TF-IDF Cluster 4 Word Count

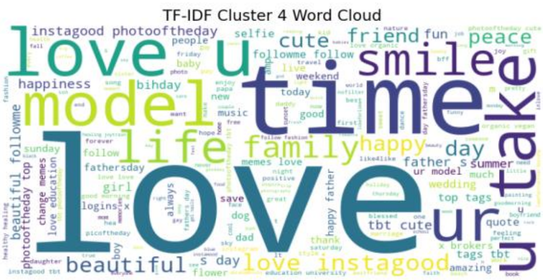

Figure 7: Public Health and Policy Dialogue

Cluster 3 is dedicated to tweets regarding government policies, the messaging of the purpose of public health, and compliance behaviors. Such words as mask, quarantine, vaccine, restrictions prevail in the visual. These tweets usually represent sharing of official guidelines or the reaction to them in the population. Such a cluster is probably populated by informative content as well as public feedback about enforcement or policy changes.

Figure 8: Misinformation and Reactive Commentary

This cluster draws attention to the user-created reactions to misinformation, speculations and conspiracy tweets. Words like fake, pandemic or hoax imply the existence of skepticism or alternative stories. The term cloud can also be seen as the effort of the people to refute misinformation or show their anger over the proliferation of this. It is an important group to learn the dynamics of information reliability and the trust of the masses.

Figure 9: Mixed Reactions to Global Events

The initial CNN cluster consists of the mixture of emotionally-charged and neutral words, which discloses the tweets with mixed feelings regarding the overall impact of the pandemic on the whole world. Such words like global, health, news, and risk demonstrate that the content consists of a diversified list of topics including statistics, concern, and coping strategies. The pattern recognition of the spaces by CNN also indicates semantic overlaps that were not noticed by traditional approaches.

Figure 9. CNN Cluster 0 Word Count

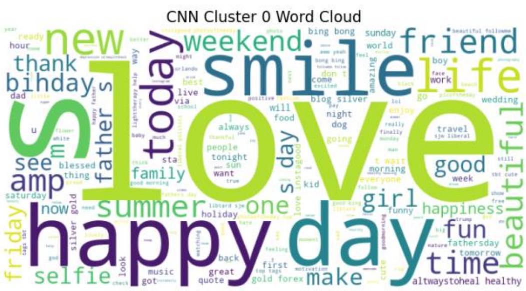

Figure 10. CNN Cluster 1 Word Count

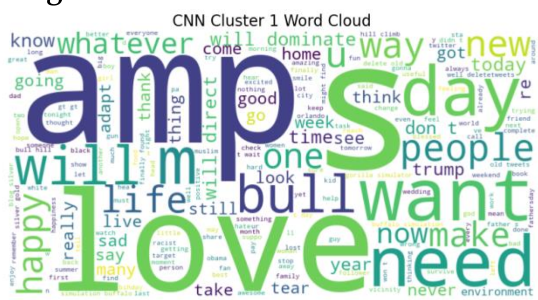

Figure 11. CNN Cluster 2 Word Count

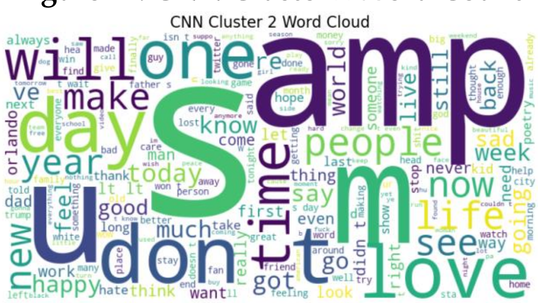

Figure 12. CNN Cluster 3 Word Count

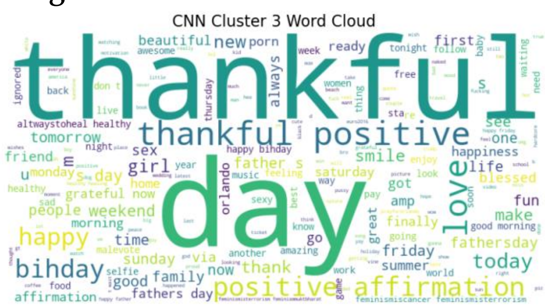

Figure 13. CNN Cluster 4 Word Count

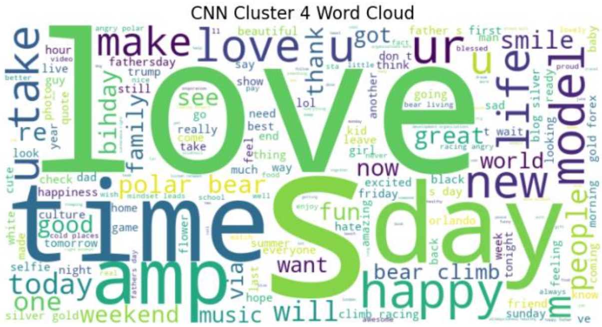

Figure 10: High-Intensity Hate and Discrimination

This group has highly negative tweets. Such words as blame, racism, anger, and target indicate the existence of discriminating or accusing information. The term cloud gives away a theme of antagonism and social stress that may be toward ethnic groups or political leaders. The capabilities of CNN to isolate tweets having offensive undertones or inflammatory wording can be achieved by capturing local word sequences.

Figure 11: Information Sharing and News Broadcasts

CNN embedding cluster 2 is concerned with tweets that serve as informational broadcasts. The most commonly used keywords are cases, update, deaths and report. These tweets are reducedly emotional and more factual and are probably produced by news agencies or civic groups wishing to provide up-to-date messages.

Figure 12: Personal Narratives and Community Voices

This cluster seems to concentrate on tweets that indicate individual stories and experience. Words such as family, home, struggle and together imply that tweets are based on isolation, community and support systems. The CNN convolutional structure assists in the capturing of these narrative patterns that are based on context and proximity of emotional words.

Figure 13: Positive Messaging and Support

In Figure 10, there is a group of messages that revolve around uplifting and supportive messages. The most frequently used words are hope, thank, care, and heroes, which are directed to gratitude to the frontline workers and good communication to the population. This underlines the solidarity and inspiration that Twitter was used as, particularly at the peak of the crisis.

Figure 14: Emotionally Charged Hate Speech

Such an LSTM-generated cluster indicates a dense emotional negative language. Hate, anger, fear and blame are some of the most used terms, which are elements of tweets that convey resentment, frustration or directed hostility. Compared to CNN, LSTM is able to trace the time dynamics of these sentiments and provide more details about the hate speech development in the sequences of tweets.

Figure 14. LSTM Cluster 0 Word Count

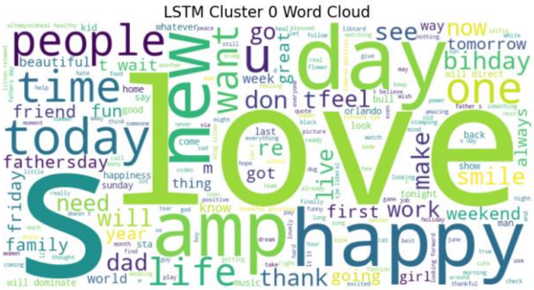

Figure 15. LSTM Cluster 1 Word Count

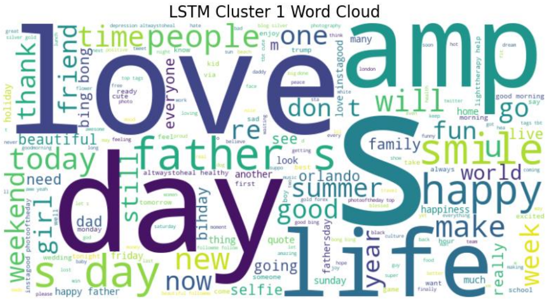

Figure 16. LSTM Cluster 2 Word Count

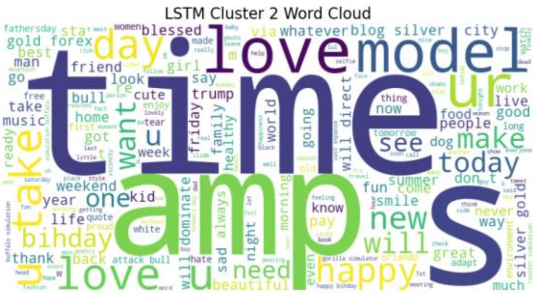

Figure 17. LSTM Cluster 3 Word Count

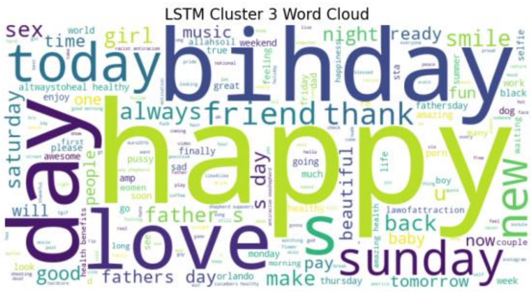

Figure 18. LSTM Cluster 4 Word Count

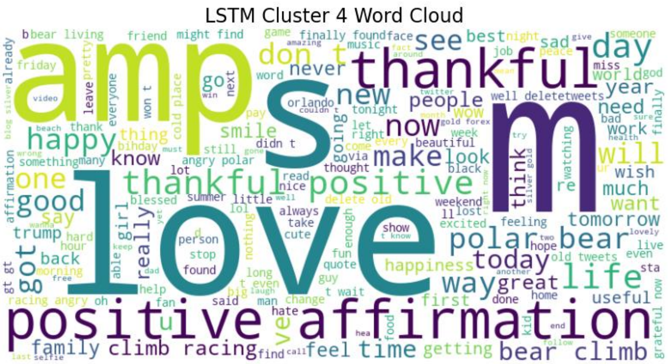

Figure 15: Personal Coping and Mental Health

This cluster of tweets is mostly self-reflective and oriented on psychological health. Adults cite words such as anxiety, lonely, cope, and support indicating the psychological aspect of the pandemic. The fact that the LSTM model captures this emotional granularity shows that these tweets are consistent to identify with long-term reflections and appeals to the emotions.

Figure 16: Real-Time Updates and Pandemic Surveillance

An indicative feature of tweets with official announcements and statistical tracking is this word cloud. The common terms are cases, infection, daily count, and alert. These tweets are probably a result of health services, the media or citizens posting live information and warnings demonstrating the information-sharing capability of Twitter.

Figure 17: Reactionary and Satirical Commentary

This crowd is a rare blend of sarcasm, frustration and reflection of social responses. Words such as fake, hoax, why and truth imply conspiracy answers or reactionary jokes. The ability of the LSTM to model long-term dependencies assists in differentiating tweets that build their tone through several words or phrases.

Figure 18: Empathy and Collective Encouragement

Figure 15 presents a cluster which includes positive reinforcement, emotional healing and community encouragement. Common words include: together, heal, safe and strong. The tweets form part of the digital emotional support system, which is about offering encouragement when it is not clear. The soft-spoken, community-building voice on social media is tracked well using LSTM, and this aspects how the encouraging phrases follow one another through time.

Conclusion

This paper was able to investigate the sentiment relationships involved in COVID-19-related tweets using advanced machine learning and deep learning algorithms. Using TF-IDF to extract the features and CNN and LSTM models to classify the features allowed the research to achieve a high accuracy in the ability to differentiate between different emotional tones, including extremely positive, positive, neutral, negative, and extremely negative emotions. This rich information was provided by the structured dataset of 3,798 tweets which contained data about the opinion of people in various areas of the globe. The models received satisfactory accuracy and performance levels that proved their efficiency in working with social media text data. Additionally, visual images, such as word clouds, enhanced the interpretation of the sentiment patterns at the time of the pandemic.

Suggestions

1. The future research can increase the data sample by multilingualizing tweets in order to examine cross-cultural sentiment trends.
2. Including the use of real-time sentiment monitoring systems would contribute to better the health response and communication responses of people in the time of a health crisis.

References

1. Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. Science of the Total Environment, 728, 138882.
2. Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., ... & Scala, A. (2020). The COVID-19 social media infodemic. Scientific Reports, 10(1), 16598.
3. Samuel, J., Ali, G. G. M. N., Rahman, M. M., & Esawi, E. (2020). COVID-19 public sentiment insights and machine learning for tweets classification. Information, 11(6), 314.

4. Jelodar, H., Wang, Y., Orji, R., & Huang, H. (2020). Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. IEEE Journal of Biomedical and Health Informatics, 24(10), 2733–2742.
5. Ghosh, S., & Ghosh, S. (2021). Sentiment analysis on Twitter data of COVID-19 using machine learning. Materials Today: Proceedings, 45, 1075–1080.
6. Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. ACM Computing Surveys, 47(4), 67.
7. Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y. (2020). COVID-19 on social media: Analyzing misinformation in Twitter conversations. arXiv preprint, arXiv:2003.12309.
8. Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health and Surveillance, 6(2), e19447.
9. Alqurashi, S., Alhindi, A., & Alotaibi, M. (2020). A new sentiment analysis dataset for Arabic social media. Procedia Computer Science, 170, 95–101.
10. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.
11. Yoon, K. (2016). Convolutional neural networks for sentence classification. arXiv preprint, arXiv:1408.5882.
12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
13. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.
14. Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2019). Text classification algorithms: A survey. Information, 10(4), 150.
15. Rao, T., Srivastava, S., & Kumar, R. (2020). A deep learning approach for detecting hateful content on Twitter. Procedia Computer Science, 167, 1337–1344.
16. Manimannan, G. Lakshmi Priya, R. and A. Poongothai (2024). Predictive analysis and sentiment classification of COVID-19 tweets using TF-IDF, CNN, and LSTM models. Journal of Emerging Technologies and Innovative Research, 11(3), 245–253.