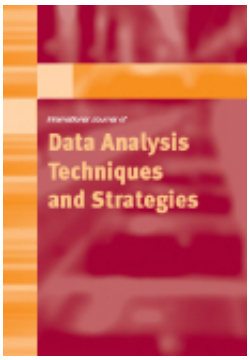


International Journal of Data Analysis Techniques and Strategies

ISSN online: 1755-8069 - ISSN print: 1755-8050
https://www.inderscience.com/ijdats

Sentiment analysis of Twitter data using machine learning: COVID-19 perspective

Shobhit Srivastava, Mrinal Kanti Sarkar, Chinmay Chakraborty

DOI: 10.1504/IJDATS.2024.10062934

Article History:

Received:	04 February 2023
Last revised:	06 February 2023
Accepted:	20 November 2023
Published online:	19 March 2024

Sentiment analysis of Twitter data using machine learning: COVID-19 perspective

Shobhit Srivastava* and Mrinal Kanti Sarkar

University of Engineering & Management, Jaipur,
Rajasthan 303807, India
Email: shobhitan@gmail.com
Email: mrinalk.sarkar@uem.edu.in
*Corresponding author

Chinmay Chakraborty

BIT Mesra,
Mesra, Jharkhand 835215, India
Email: cchakrabarty@bitmesra.ac.in

Abstract: The 2019 COVID-19 pandemic has affected people worldwide. Social media has become a global platform for individuals to voice their diverse perspectives on the pandemic, which has significantly altered their lives during and beyond lockdown periods. Twitter, a leading social media platform, experienced a surge in coronavirus-related tweets encompassing a spectrum of positive, negative and neutral opinions. Coronavirus transmits between humans in numerous ways. It irritates the lungs. This makes Twitter a perfect platform for expressing opinions. Twitter data from across the world was collected and analysed for sentiment in order to better understand public opinion and prepare for COVID-19 (Tusar et al., 2022). In this article, our aim is to compare the neural network techniques and indicate the share of their performance measures. We use kNN and neural network algorithms for these and use the MSE factor as a key of comparison. However, we use other performance measures too for better analysis of the result. Our main focus in this study is to analyse the performance partition of the kNN algorithms, including the performance portion of the each algorithm.

Keywords: COVID-19; social media; sentiment analysis; Twitter; machine learning; neural networks; KNN; neural network.

Reference to this paper should be made as follows: Srivastava, S., Sarkar, M.K. and Chakraborty, C. (2024) 'Sentiment analysis of Twitter data using machine learning: COVID-19 perspective', *Int. J. Data Analysis Techniques and Strategies*, Vol. 16, No. 1, pp.1–16.

Biographical notes: Shobhit Srivastava is a dedicated researcher currently immersed in the pursuit of a PhD in Computer Science and Engineering at the University of Engineering and Management, Jaipur, Rajasthan, India. His research focus study is machine learning techniques for behavioural analysis, showcasing his commitment to advancing the understanding and application of intelligent systems.

Mrinal Kanti Sarkar is a Professor and the Head of Computer Science and Engineering Department, University of Engineering and Management, Jaipur, Rajasthan, India. He published more than 20 research papers in various fields of computer science.

Chinmay Chakraborty is an Assistant Professor at Birla Institute of Technology, Mesra, India. He is a Post-doc Fellow at the Federal University of Piauí, Brazil, and visited University of Malta, Europe. His main research interests include the internet of medical things (IoMT), AI/ML, communication and computing, Telemedicine, m-Health/e-health, and medical imaging. He has widely published 200+ articles in peer reviewed international journals, conferences, book chapters, 25+ books, 6+ patents, and 20+ special issues in the field (Google h-index 40/ i10-index 119, Scopus h-index 35, ISI-WoS h-index 26).

1 Introduction

After the virus had already spread globally, the 2019 coronavirus disease (COVID-19) outbreak was officially recognised in Wuhan, China. The World Health Organization (WHO) designated it as a pandemic, acknowledging its widespread and severe impact. There are currently a significant number of people all around the world who are afflicted by this disease. Presently, COVID-19 presents a considerable threat to global populations, especially in areas experiencing a surge in pneumonia-like symptoms. Its numerous effects on the human body, ranging from severe respiratory distress to multi-organ failure, can lead to rapid fatality. Twitter's importance has skyrocketed in recent months as a result of the ongoing COVID-19 conflict around the world, which has resulted in the detention of the vast majority of people media platforms, such as Twitter, have consistently played a crucial role in facilitating communication, self-expression, and information sharing during various disasters, such as cyclones, Ebola, floods and Zika (Manguri et al., 2020). Amidst the isolating effects of the pandemic, social media has provided a valuable space for individuals to voice their emotions and connect with others. While social media can serve as a source of real-time and essential information about COVID-19, it is essential to remain vigilant about the potential for inaccurate or misleading content. If someone finds incorrect or sad material on social media, their already difficult situation may become even worse. The rise of the internet has ushered in a new era of communication, fundamentally altering the way we express our thoughts, opinions and experiences. No longer confined to traditional channels like letters, phone calls, or face-to-face conversations, individuals now engage in a vibrant digital discourse through a diverse array of online platforms (Pokharel, 2020). At the forefront of this transformation are social media giants like Facebook, Twitter and Google Plus, which have become virtual hubs for millions of users to share their perspectives, connect with others, and engage in real-time conversations (Kaila and Prasad, 2020). These platforms have democratised self-expression, empowering individuals to participate in a global dialogue and contribute to the ever-evolving tapestry of human experiences. The sheer volume and diversity of user-generated content on social media platforms have created a rich repository of data that provides valuable insights into human behaviour, societal trends and global events. Tweets, status updates, blog posts, comments, reviews, and other forms of social media expression offer a real-time snapshot of public sentiment, allowing us to gauge the pulse of society and track the evolution of public opinion (Awotunde et al., 2022). In addition to this, social media gives companies the option to connect with their customers for advertising by providing a platform on which to do so (Ramteke et al., 2016). When it comes to making decisions, most people put a significant

amount of stock in the user-generated content that can be found online. For example, before choosing the purchase of a product or the utilisation of any service, a person will first seek evaluations of that product or service online and engage in conversation regarding that product or service on social media (Dubey, 2020). The vast ocean of user-generated content could easily drown an ordinary user, hindering their ability to fully comprehend the intricate and multifaceted nature of the information landscape (Shah and Swaminarayan, 2022). As a result of the necessity of automating this process, numerous systems for analysing sentiment are currently in widespread use (Rezaee et al., 2023).

2 Review of literature

A lot of academics have been hard at work performing sentiment analysis on data from a variety of social media platforms, most notably Twitter. These scholars have produced a lot of important discoveries that assist in identifying the attitudes or feelings of users in a range of different scenarios while pandemics are occurring throughout the world. This section discusses a selection of the significant articles that served as references throughout the project. Following the topic of public concern when an epidemic is occurring is an essential component of public health and should be considered a top priority (Shin et al., 2016). Leveraging real-time social media platforms can provide valuable insights into public health issues and community sentiments, particularly in the context of outbreaks. Pokharel (2020) researched COVID-19 outbreaks in Nepal and analysed sentiments gleaned from Twitter data. The data for this study was collected from Twitter users who self-identified as being located in Nepal between May 21, 2020 and May 31, 2020. This was done using the Twitter API and the Tweepy Python package. The Twitter API is a set of tools that allows developers to access and interact with Twitter data. Tweepy is a Python library that makes it easier to use the Twitter API. By using the Twitter API and Tweepy, researchers can collect data from Twitter for a variety of purposes, including studying public opinion, tracking trends, and analysing social media sentiment. In this case, the data was collected to study public health issues and community ideas in Nepal during the COVID-19 pandemic. The 615 tweets that were obtained were analysed with TextBlob Library, which is one of the Python sentiment analysis approaches. Kaila et al. (2020) employed Portuguese-language tweets to delve into the evolution of COVID-19-related discourses in the context of escalating political tensions between Brazil and China. By analysing the content of these tweets, the researchers sought to illuminate the intricate interplay between public health concerns and geopolitical dynamics. Their findings highlight the profound influence of political factors on shaping public perceptions and narratives surrounding a global pandemic. These tensions were caused by a trade dispute between the two countries. Between March 19 and April 1, 2020, they used 1.6 million tweets altogether. Following the completion of the various filtering processes, this dataset was then subjected to thematic and sentiment analysis. The tweets that were gathered for this research were used. Ramteke et al. (2016) acquired 3,000 English tweets from a variety of nations, and following the pretreatment phase of these tweets, did an emotional analysis on the remaining 2,058 tweets. Shin et al. (2016) examines the tweets that were collected on Twitter during the first few months of the COVID-19 epidemic in Europe and evaluate the sentiments included within them in their articles. The datasets contain 4.6 million

geo-tagged Twitter messages that were collected between the months of December 2019 and April 2020. The bag-of-words method was utilised to do sentiment analysis by Turney (2002). This method represents a document as nothing more than a collection of words and it does not take into account the connections that exist between individual words. To identify the attitude conveyed by the entire manuscript, the attitudes conveyed by each word were analysed, and then the results of those analyses were aggregated using several functions. The lexical database WordNet was utilised by Kamps et al. (2004) to determine the emotional connotations associated with a word along a variety of parameters. They assessed the semantic polarity of adjectives and established a distance metric based on WorldNet. Table 1 illustrates the potential contribution of the researchers.

Table 1 Potential work of authors

Ref.	*The behaviour of the dataset used*	*F1 score*
D'Andrea et al. (2019)	Twitter data based on vaccination	F1 score: 67.4%
Chatsiou (2020)	Sentences from the media on COVID-19	F1 score 70.65%
Jelodar et al. (2020)	Multi-class Twitter dataset on COVID -19	F1 score 83.15%
Garcia and Berton (2021)	Bi-class dataset collected using Twitter	F1 score 82.3%
Naseem et al. (2021)	COVIDSENTI dataset	F1 score 93.8%
Sitaula et al. (2021)	Dataset of Twitter data of Nepal	F1 score 69.4%
Shahi et al. (2022)	Feature-tuned dataset collected from Twitter	F1 score 73.6%
Sitaula and Shahi (2022)	Multi-class Twitter dataset on NepCOV19	F1 score 73.3%
Saadah et al. (2022)	Dataset of Twitter data of Indonesia on vaccination	F1 score 82%
Srivastava et al. (2023)	Linguistically categorised and trained dataset	F1 score 93%

3 Sentiment analysis

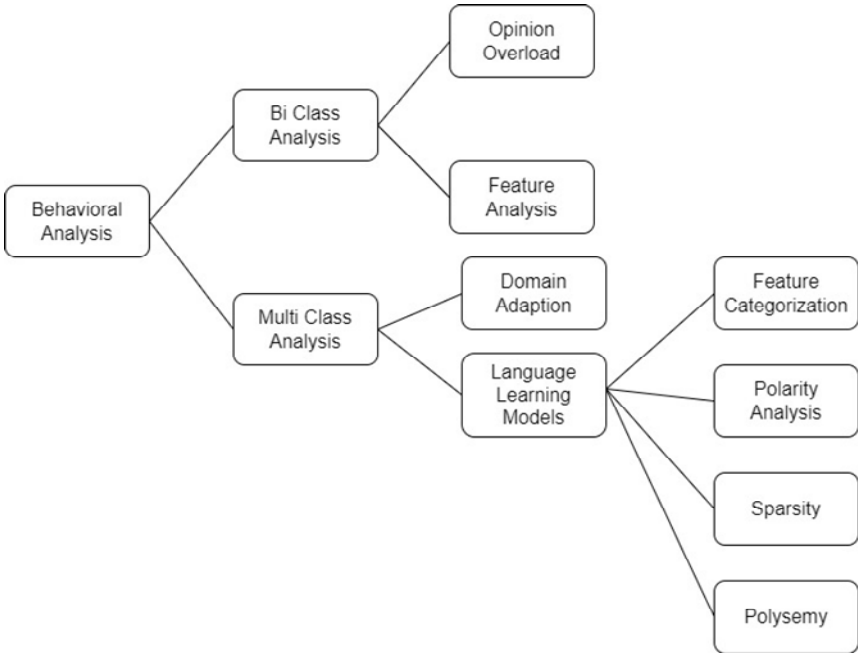
When the data was collected, it was loaded into Orange Data Mining Software to visualise the sentiments conveyed in the tweets. The corpus was prepared using Orange, where standard data preprocessing methods including tokenisation, transformation, and filtering were applied. Positive sentiment was shown by a score of more than 0.05, a neutral mood by a score between –0.05 and 0.05, and a negative sentiment by a score of less than 0.05. Sentiment analysis is based on the idea that the language we use can reveal our underlying attitudes and emotions. Text sentiment classification employs diverse methods to recognise and categorise emotions within textual content. These techniques include:

- Lexicon-based methods: In these approaches, a lexicon comprising words and phrases linked to positive, negative, or neutral sentiments is employed (Rose et al., 2018). The sentiment orientation of a given text is gauged based on the quantitative analysis of positive, negative, and neutral words and expressions it encompasses.

- Machine learning methods: In these methodologies, algorithms are employed to acquire the proficiency of sentiment classification in textual content. These algorithms undergo training using an extensive body of text meticulously annotated with positive, negative or neutral labels.
- Hybrid methods: These approaches integrate both lexicon-based and machine learning methodologies to conduct sentiment analysis. However, the task is not devoid of challenges. An obstacle lies in accurately classifying sentiment in text characterised by sarcasm, irony, or ambiguity, where the conventional methods may encounter difficulty. Another challenge is that sentiment analysis is often language-specific. This means that algorithms that are trained on English text may not perform well when applied to text in other languages.

Despite these challenges, sentiment analysis is a valuable tool that can be used to gain insights into public opinion and customer behaviour. It is a powerful tool that can be used to understand the world around us.

Figure 1 Sentiment analysis domains



3.1 Twitter (X)

Twitter is a popular real-time micro-blogging website that enables users to communicate brief bits of information in the form of 'tweets', each of which can only be up to 280 characters long. Tweets are short messages that users post to convey their thoughts on a variety of issues that are relevant to their lives. Twitter is an excellent tool for gathering the general public's opinion on a variety of topics for research purposes. Tweets make up the sentiment analysis, opinion mining, or natural language processing

corpus (Tyagi and Tripathi, 2019). Social media platforms host this collection. Twitter has transformed into an indispensable tool for companies to monitor their reputation and brand perception by extracting and analysing public tweets about their products, services and competitors. Twitter receives one million messages daily from 500 million users. Sentiment analysis is the fastest, most complete, and easiest on the internet. Social media generates opinions, and with the tremendous rise of the internet, super volumes of opinion writings are available for examination. Tweets, reviews, blogs, and discussion forums can contain these texts.

4 Problem statement

The purpose of this study is to analyse how people are feeling at specific points in time and capture their processes through analysis of tweets made on Twitter. Consequently, the following issues would be the primary focus of the research: leveraging the RTweet package in the R programming language and establishing a connection with the Twitter API, to gather tweets of interest. Subsequently, conduct preliminary data cleaning on the collected tweets to eliminate extraneous elements such as white spaces, links, punctuation marks, stop words and retweets. Perform an analysis of the output after calculating the sentiment using the syuzhet package. Sentiment analysis has been performed on the tweets that have been posted in English to gain a better understanding of how people from the various infected countries have reacted throughout this pandemic emergency to cope with it. The sentiment analysis will be carried out by using, preprocessing, and applying text mining algorithms to the collected tweets as the data source. Our analysis aims to differentiate the performance of kNN algorithms categorically.

5 Methodology

Following the phase of gathering the data, the first stage is the phase of preparing the data. As a preliminary step, tweets contained in fields like ‘username tweet’ and ‘ID’ were excluded since they held no relevance to the feature selection process and the subsequent data analysis. The dataset has been cleaned of all fields except the ‘text’ and ‘date time’ fields, which are going to be used for this investigation. Following that, all of the capital letters were changed to their lowercase counterparts, and string expressions were applied to all of the numeric characters that comprised the text. Because there were far too many instances of identical tweets included in the dataset, we had to get rid of some of them. The work steps are depicted in Figure 2.

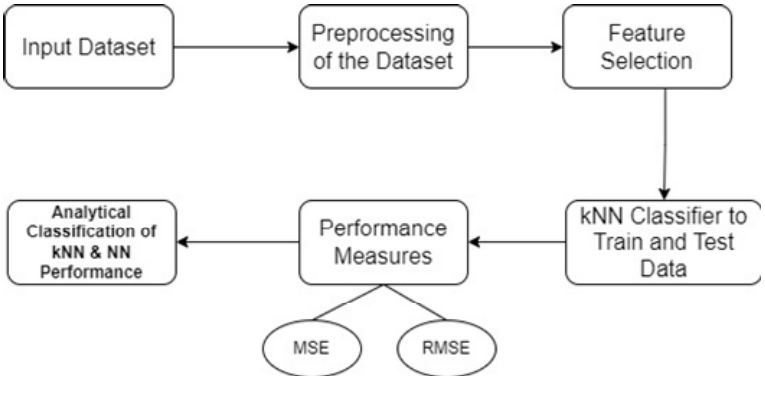
Word clouds are helpful tools for providing a visual summary of extensive volumes of text data. In this investigation (Figures 3 and 5), the word cloud library for Python was utilised to visualise the text that was collected the most frequently. Data have been collected from Kaggle. Within the Python programming environment, a multitude of packages furnish a collection of functions specifically crafted for text classification and sentiment analysis. Some of the programmes that are utilised are TM, tidytext, wordcloud, dplyr, and syuzhet. R’s tm package is used by applications that perform text mining. The goal of using Tidytext is to transform unstructured text data into a form that is more amenable to analysis. The Wordcloud package includes various elements that can be utilised in the construction of attractive word clouds. Dplyr, a data manipulation

grammar, provides a streamlined approach to conquering common data wrangling hurdles with a standardised collection of verbs. Feelings and sentiment dictionaries can be retrieved from the syuzhet package, which also contains plots that are produced from sentiment. Imported are datasets obtained from the neighbourhood library. The assembled dataset will be housed in a corpus variable, facilitating preprocessing using Python-based software. This preprocessing stage entails data cleansing and preparation for subsequent analysis or modelling. The algorithm of the workflow is described at Algorithm 1.

Algorithm 1 Hybrid kNN algorithm

Function – Hybrid_kNN-LR (COVID-19 Twitter dataset)
Input: textual dataset
Output: predicted class labels for the test dataset
For each text ti do
Preprocessed text (Pt) ← preprocessing of text (ti)
Word Tokens (Wtokens) ← tokenisation of Pt
Redefined text (RT) ← feature vector (Pt)
Ready to use text (RU) ← padding (RX)
kNN (k) ← kNN_Uncased_Model (RU)
NN (n) ← NN_Uncased_Model (RU)
Output (O) ← deep (FC)
Determine the MSE of kNN k(M)
Determine the MSE of NN n(M)
Show classification result (k(M) and n(M))
End For

Figure 2 Methodology steps



5.1 Dataset used

Dataset has been collected from Kaggle and the world cloud and dataset structure is described under Figures 3 and 4 and in Table 2 depicts the sample of the dataset. The

dataset having the fields like tweet Id, tweet date, location, original tweet, sentiment, etc. In this study, we use only location, original tweet and sentiment as the key feature.

Figure 3 World cloud of dataset with location feature (see online version for colours)

Figure 4 World cloud of dataset tweet feature (see online version for colours)

Table 2 Overview of the dataset

Location	*Original tweet*	*Sentiment polarity*
India	My next one who is more important for you right now from communications perspective in order of priority consumer employee or govt. authorities husain Ray C 19.	Positive
England, UK	Are recruiting support volunteers to help people in need during this unprecedented time tasks could include reaching out with information dog walking calling lonely people picking up shopping or posting mail sign up online.	Positive
Sheffield	With the ongoing situation in regards to #Covid-19, more and more people are using online shopping. Amazon have set up #AmazonSmile so that while you do your online shopping, you can support a charity of your choice! Find out more here.	Extremely positive
Kigali, Rwanda 2020	Drones have been implemented in ChinaÂ's battle against the #covid19 outbreak. #UAS #tech was used to spray disinfectant in public spaces, to deliver medical samples and to deliver consumer goods to their citizens. #coronavirus #pandemic #techforgood.	Negative
Pittsburgh, Glasgow, Lima	What is there for travel #ecommerce players to learn from fraud perspective as online shopping witnesses a surge owing to the #coronavirus pandemic? @Riskified #payments #fraudprevention #dataanalytics #giftcards #machinelearning #airlines https://t.co/YTDVD0brj1.	Negative
Indore, India	1 in 5 rated APAC companies have high exposure to disruptions and are sensitive to shifting consumer demand and travel restrictions Another 36 has a moderate potential for implications to their credit quality or ratings.	Negative
Dubai	Benchmark #Brent crude oil futures rose as high as $33.37 a barrel on rising hopes of a new global deal to cut global crude supply.	Extremely negative
Houston	Asia petrochemical shares mixed on virus fears; oil reverses early losses #ICIS #coronavirus #COVID19 #Asia #petrochemicals #crude #oil #prices #supply #energy #chemicals https://t.co/KTCG0VXNCi.	Negative
Houston	Europe chem prices crash, 32% of refinery capacity restricted, offline #ICIS #coronavirus #COVID19 #Europe #chemicals #prices #refinery #oil #ethylene #propylene #benzene https://t.co/Xk4FK0CPwh.	Extremely negative
Houston	US-listed shares of chemical companies fell even as oil prices rose for the second day on the prospect that the world's major oil producers could reach a deal to limit output. #petchemindustry #petrochemicals #stockmarket #oilcrash #covid19 #coronavirus.	Positive

5.2 kNN

k-nearest neighbours algorithm is a simple, versatile, and non-parametric supervised learning algorithm used for both classification and regression tasks. The algorithm is based on the simple idea that similar things are likely to be together. In other words, the algorithm assumes that the data points that are closest to each other in terms of some distance metric are likely to have the same label. To classify a new data point, the KNN algorithm first calculates the distance between the new data point and all of the data points in the training set. Then, it selects the k data points that are closest to the new data point. Finally, it assigns the new data point the label that is most common among the k-nearest neighbours. The choice of the value of k is important. If k is too small, then the algorithm may be oversensitive to noise in the data. If k is too large, then the algorithm may not be able to capture the local patterns in the data.

Test point: X.

Define the set of k-nearest neighbours of X as S_X. Formally S_X is defined as $S_x \subseteq D$ s.t. $|S_x| = k$ and $\forall (x', y') \in D \setminus S_x$.

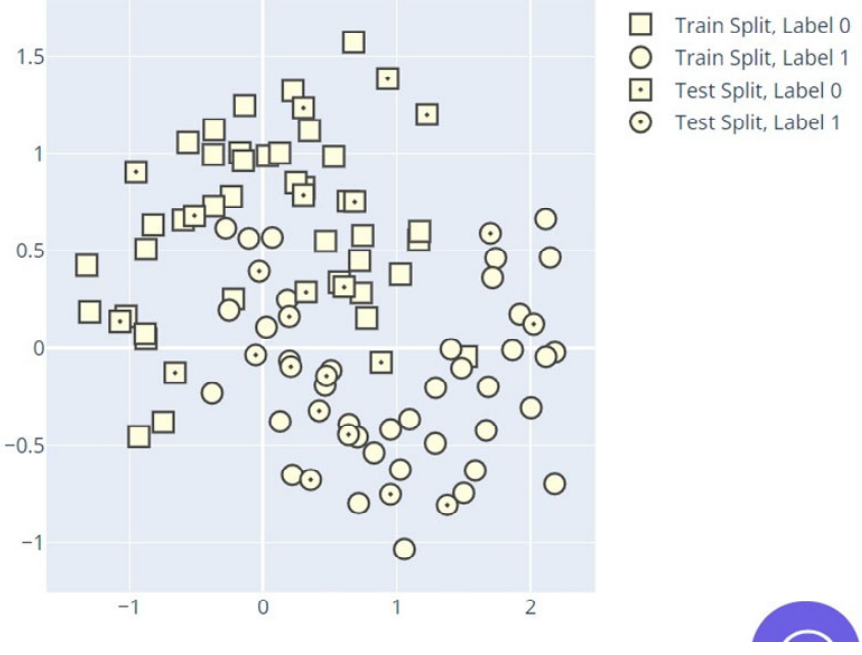
$$\text{dist}(x, x') \geq \max(x'', y'') \in S_x \text{dist}(x, x'') \tag{1}$$

(i.e., every point in D but not in S_x is at least as far away from x as the furthest point in S_x). We can then define the classifier h() as a function returning the most common label in S_x:

$$h(x) = \text{mode}\left(\left\{y'' : (x'', y'') \in S_x\right\}\right) \tag{2}$$

where mode(·) means to select the label of the highest occurrence.

Figure 5 KNN classification (see online version for colours)



The k-nearest neighbour classifier fundamentally relies on a distance metric. The better that metric reflects label similarity, the better the classified will be. The most common choice is the Minkowski distance.

$$\text{dist}(x, z) = \left(\sum r = 1d|Xr - Zr|p\right)1/p \tag{3}$$

6 Implementation and analysis

It is necessary to preprocess every piece of text before continuing with the processing of the text. Several text preparation strategies were utilised so that unused document type entry datasets may be deleted. When it comes to text preprocessing, many different strategies can be used; however, only a few of them have been utilised in this study as models. Characters with special meanings, such as @, #, and /, do not add anything to the overall meaning of the review. In Python programming, the type of data known as the term document matrix is frequently put to use. The input makes primary use of this to determine the frequency of individual words. The corpus variable can be transformed into a matrix structure, allowing for efficient text representation and analysis. Furthermore, a word cloud is generated based on the selected dataset to improve the visualisation. This word cloud only displays the most frequently occurring words, highlighting the prominent themes and topics within the text. In this study, we focus on analysing two separate datasets (Figures 4 and 6). The following is a description of these two datasets: Dataset-I. Dataset-I, sourced from Kaggle, features a substantial collection of tweet texts pertaining to COVID-19, incorporating keywords like 'Corona', 'Covid-19', and 'Coronavirus' (case-insensitive). Given Twitter's status as the most popular social media platform in the USA, our study concentrated on these nine states (Sahayak et al., 2015). Users in these states contribute to a substantial share of tweets posted on Twitter. To explore the potential connection between tweet volume and disease incidence, a second dataset was acquired from GitHub, containing COVID-19 case numbers. This was done in Dataset-II. It scans massive databases for intriguing patterns. It stems and removes stop words to do this. This study compares machine learning methods, specifically kNN algorithms, with respect to MSE, accuracy, F1 score and others. This study focuses that if we apply kNN algorithms to datasets, which one performs better. We hope this study will help text mining researchers grasp the multiple preprocessing options.

Figure 6 is showing the complete description of the workflow for the evaluation of different models concerning sentiment analysis. The evaluation must be most important for the correct accuracy level and fill the eject gap in form of values between the data. Figures 7–12 are showing the sentiment analysis criteria for COVID-19 variances and also show the measurable structure of variances. Tables 3 to 5 are given the model accuracy, precision, recall, and F1 measurable values which denote the difference between 100% accuracy levels.

Figures 10 and 11 show the confusion matrix from the k-nearest neighbours clustering with k = 4 for all 34 attributes and all eight classes, with the true class labels on the x-axis and the class predictions on the y-axis. Correct categories are on the first diagonal. The bottom right cell reflects accuracy overall.

Figure 6 Work flow for evaluations

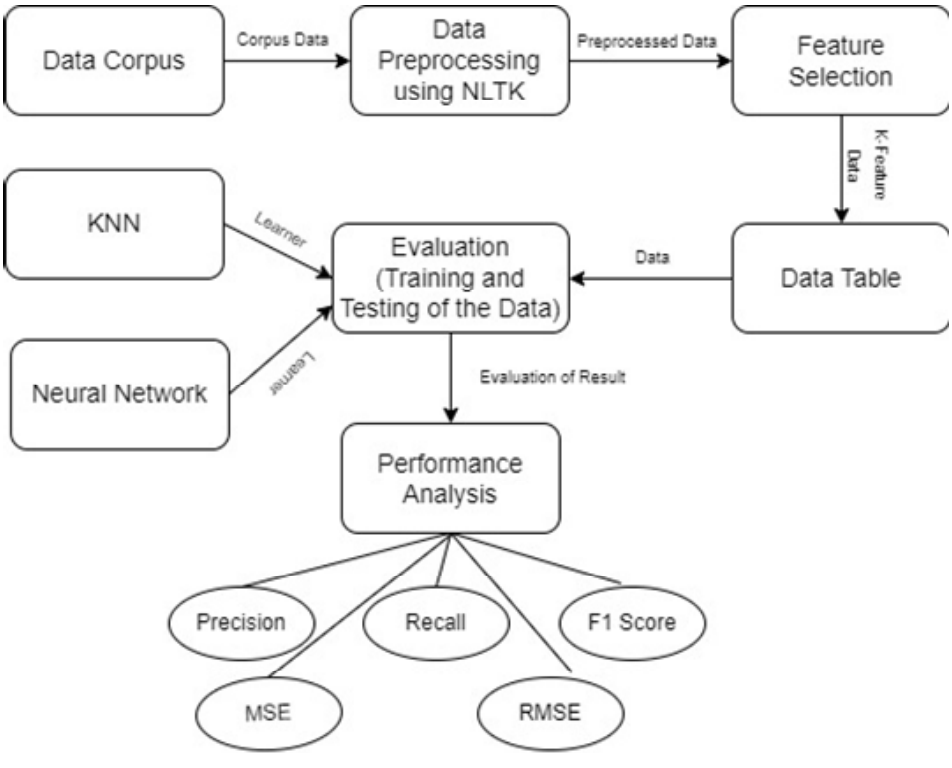


Table 3 Performance comparison 1

Algorithm	*Precision*	*Recall*	*F-score*
KNN	0.028	0.999	0.995
Neural network	0.972	0.001	0.005

Table 4 Performance comparison 2

Test and score of data table_1			
	RMSE	*MSE*	*R2*
kNN	1.72	1.76	–0.036
Neural network	0.458	0.242	–0.69

Table 5 Performance comparison 3

Test and score of data table_2			
	RMSE	*MSE*	*R2*
kNN	1.56	1.056	–0.021
Neural network	0.408	0.096	–0.003

Figure 7 Complete evaluations

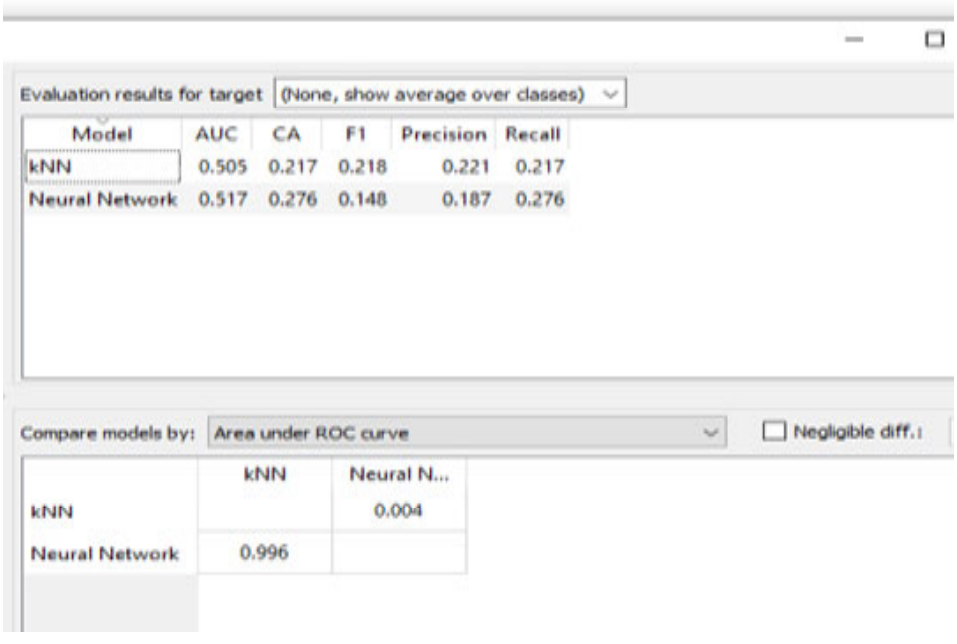


Figure 8 Precision evaluation

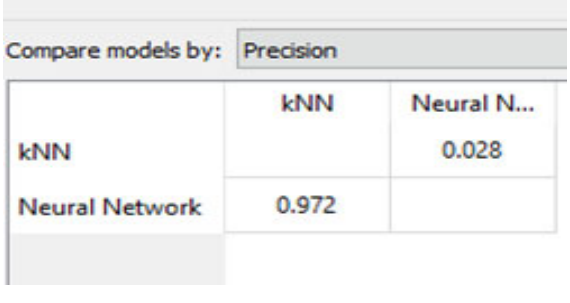


Figure 9 Recall evaluation

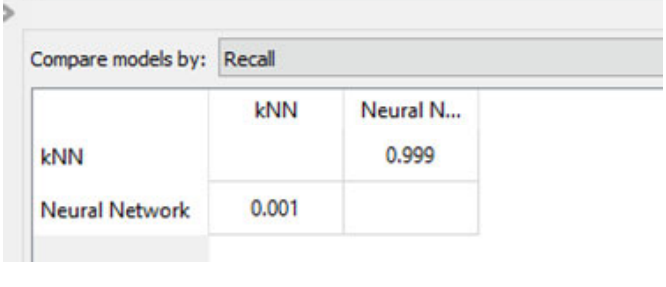


Figure 10 F1 evaluations

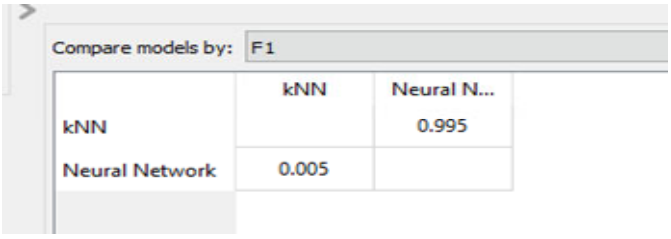


Figure 11 Confusion Metrices_kNN (see online version for colours)

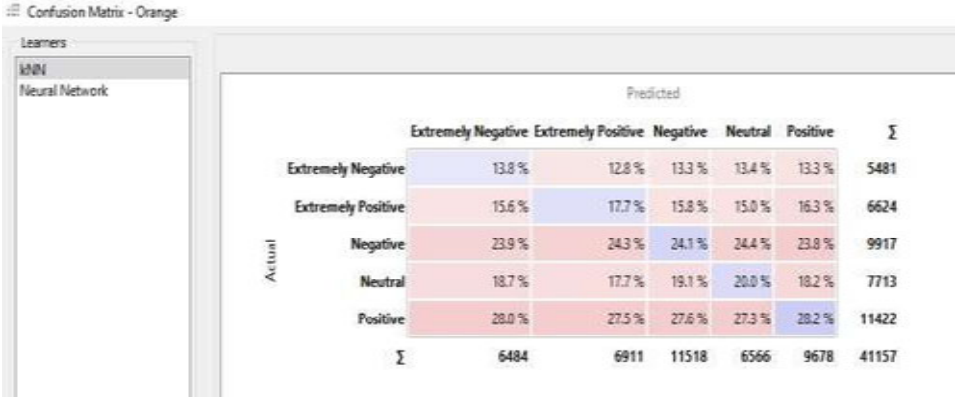
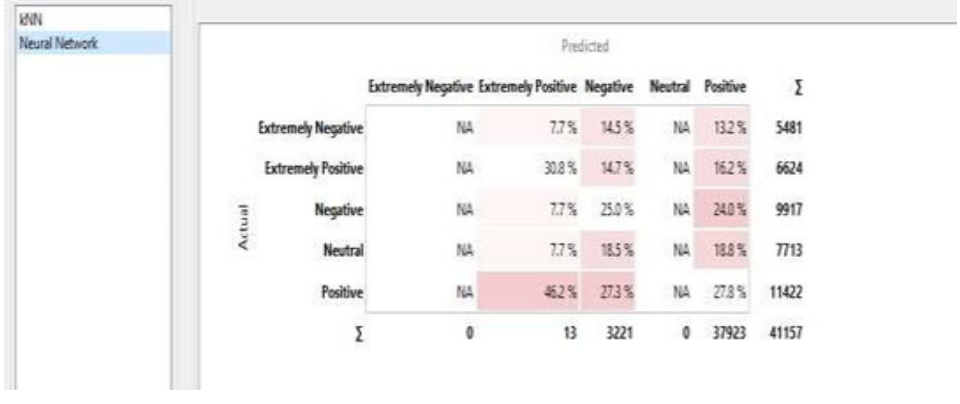


Figure 12 Confusion Metrices_Neural network (see online version for colours)



7 Conclusions

This study, which set out to examine how people felt during the COVID-19 pandemic, has been carried out successfully. Several methods for gathering public sentiment are compared and contrasted in this investigation. Some of these instruments include machine learning and lexicon-based approaches, as well as cross-domain and

cross-lingual techniques, and a few assessment measures. The findings of this study imply that machine learning approaches such as kNN and neural networks are the most accurate and should be utilised as the baseline learning methods in human-labelled texts due to their low labour requirements. Also, we looked at how various characteristics affected the classifier. With cleaner data, we can conclude that our findings will be more reliable. Research results will be utilised to learn how Indian residents generally feel about the COVID-19 vaccination. It is hoped that this would help public health professionals better comprehend the favourable, negative, and neutral sentiments expressed on Twitter about the COVID-19 vaccination. Health officials and administrators may find the data helpful in their efforts to educate the public about the benefits and hazards of the COVID-19 vaccination. The research may help the healthcare sector mitigate the effect of dissuasive messages and amplify the effect of encouraging ones to boost vaccination rates. An understanding of how online studies linked to social media data may be developed for information extraction and analysis is crucial for gaining intriguing insights into COVID-19 vaccinations and other scenarios.

References

Awotunde, J.B., Oluwabukonla, S., Chakraborty, C., Bhoi, A.K. and Ajamu, G.J. (2022) 'Application of artificial intelligence and big data for fighting COVID-19 pandemic', in Hassan, S.A., Mohamed, A.W. and Alnowibet, K.A. (Eds.): *Decision Sciences for COVID-19. International Series in Operations Research & Management Science*, Vol. 320, Springer, Cham, https://doi.org/10.1007/978-3-030-87019-5_1.

Baccouch, C., Bahhar, C., Chakrabarty, C., Sakli, H. and Aguili, T. (2022) 'E-health system for automatic control of travel certificates and monitoring of the spread of COVID-19 in Tunisia', in Chakraborty, C. and Khosravi, M.R. (Eds.): *Intelligent Healthcare*, June, pp.479–498, Springer, Singapore [online] https://doi.org/10.1007/978-981-16-8150-9_21.

Chatsiou, K. (2020) *Text Classification of COVID-19 Press Briefings using BERT and Convolutional Neural Networks*, arXiv.

D'Andrea, E., Ducange, P., Bechini, A., Renda, A. and Marcelloni, F. (2019) 'Monitoring the public opinion about the vaccination topic from tweets analysis', *Expert Syst. Appl*., Vol. 116, pp.209–226.

Dubey, A.D.J.A.a.S. (2020) *Twitter Sentiment Analysis During COVID19 Outbreak*, 9 April, SSRN [online] https://ssrn.com/abstract=3572023 or http://dx.doi.org/10.2139/ssrn.3572023.

Garcia, K. and Berton, L. (2021) 'Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA', *Applied Soft Computing*, Vol. 101, p.107057.

Jelodar, H., Wang, Y., Orji, R. and Huang, S. (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*, Vol. 24, No. 10, pp.2733–2742.

Kaila, D.P. and Prasad, D.A. (2020) 'Informational flow on Twitter – Coronavirus outbreak – topic modeling approach', *International Journal of Advanced Research in Engineering and Technology (IJARET)*, Vol. 11, No. 3, pp.128–134, SSRN [online] https://ssrn.com/abstract=3565169.

Kaila, P., Prasad, R. and Krishna, A.V. (2020) 'Informational flow on twitter – corona virus outbreak – topic modelling approach (March 31, 2020)', *International Journal of Advanced Research in Engineering and Technology (IJARET)*, Vol. 11, No. 3, pp.128–134, SSRN [online] https://ssrn.com/abstract=3565169.

Kamps, J., Marx, M., Mokken, R.J. and De Rijke, M. (2004) 'Using WordNet to measure semantic orientations of adjectives', *InLrec 2004*, 26 May, Vol. 4, pp.1115–1118.

Manguri, K.H., Ramadhan, R.N.A. and Mohammed, P.R. (2020) 'Twitter sentiment analysis on worldwide COVID-19 outbreaks', *Kurdistan Journal of Applied Research*, pp.54–65.

Naseem, U., Razzak, I., Khushi, M., Eklund, P.W. and Kim, J. (2021) 'COVIDSenti: a large-scale benchmark Twitter data set for COVID-19 sentiment analysis', *IEEE Transactions on Computational Social Systems*, Vol. 8, No. 4, pp.1003–1015.

Pokharel, B.P. (2020) *Twitter Sentiment Analysis During COVID-19 Outbreak in Nepal*, SSRN [online] https://ssrn.com/abstract=3624719; http://dx.doi.org/10.2139/ssrn.3624719.

Ramteke, J., Shah, S., Godhia, D. and Shaikh, A. (2016) 'Election result prediction using Twitter sentiment analysis', in *2016 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, August, No. 1, pp.1–5.

Rezaee, K., Zadeh, H.G., Chakraborty, C., Khosravi, M.R. and Jeon, G. (2023) 'Smart visual sensing for overcrowding in COVID-19 infected cities using modified deep transfer learning', in *IEEE Transactions on Industrial Informatics*, January, Vol. 19, No. 1, pp.813–820, DOI: 10.1109/TII.2022.3174160.

Rose, S.L., Venkatesan, R., Pasupathy, G. and Swaradh, P. (2018) 'A lexicon-based term weighting scheme for emotion identification of tweets', *IJDATS*, DOI: 10.1504/IJDATS.2018.095216.

Saadah, S., Auditama, K.M., Fattahila, A.A., Amorokhman, F.I., Aditsania, A. and Rohmawati, A.A. (2022) 'Implementation of BERT, IndoBERT, and CNN-LSTM in classifying public opinion about COVID-19 vaccine in Indonesia', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 6, No. 4, pp.648–655.

Sahayak, V., Shete, V. and Pathan, A. (2015) 'Sentiment analysis on Twitter data', *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, Vol. 2, No. 1, pp.178–183, ISSN: 2349-2163.

Shah, P. and Swaminarayan, P. (2022) 'Machine learning-based sentiment analysis of Gujarati reviews', *IJDATS*, DOI: 10.1504/IJDATS.2022.124763.

Shahi, T., Sitaula, C. and Paudel, N. (2022) 'A hybrid feature extraction method for Nepali COVID-19-related tweets classification', *Computational Intelligence and Neuroscience*, Vol. 2022.

Shin, S-Y., Seo, D-W., An, J., Kwak, H., Kim, S-H., Gwack, J. and Jo, M-W. (2016) 'High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea', *Scientific Reports*, Vol. 6, No. 1, p.32920.

Sitaula, C. and Shahi, T.B. (2022) *Multi-channel CNN to Classify Nepali Covid-19 Related Tweets using Hybrid Features*, arXiv preprint arXiv:2203.10286.

Sitaula, C., Basnet, A., Mainali, A. and Shahi, T.B.(2021) 'Deep learning-based methods for sentiment analysis on Nepali covid-19-related tweets', *Computational Intelligence and Neuroscience*, Vol. 2021.

Srivastava, S., Chakraborty, C. and Sarkar, M.K. (2023) 'A graph neural network-based machine learning model for sentiment polarity and behavior identification of COVID patients', *Int. J. Data Sci. Anal*., https://doi.org/10.1007/s41060-023-00469-7.

Turney, P.D. (2002) 'Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews', in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, pp.417–424.

Tusar, K.D., Chinmay, C., Satyajit, M. and Ganapati, P. (2022) 'Mitigating information interruptions by COVID masks: 3-stage speech enhancement scheme', *IEEE Transactions on Computational Social Systems*, pp.1–10, DOI: 10.1109/TCSS.2022.3210988.

Tyagi, P. and Tripathi, R.C. (2019) 'A review towards the sentiment analysis techniques for the analysis of twitter data', in *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, February.