International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Review on Evaluation of Different Models for Classifying Sentiments from Twitter: Challenges and Applications

Putta Durga[1], Deepthi Godavarthi*[2]

Submitted: 21/08/2023 **Revised**: 10/10/2023 **Accepted**: 22/10/2023

Abstract: Twitter has exploded in popularity in recent years as a place for users to discuss and share their thoughts on a wide variety of products and services. Researchers in the field of sentiment analysis have taken a keen interest in Twitter because of its rapid ascent to the top of the social media platform rankings. Businesses can learn a lot about their consumers' wants and needs and whether or not their products meet those wants and needs by conducting Sentiment Analysis. It's also used in healthcare to gauge public opinion on a drug or in politics to foretell the outcome of an election. Using NLP and ML techniques, Twitter Sentiment Analysis (TSA) is a helpful method for analyzing and categorizing the tone of tweets. Twitter's real-time data and potential uses in fields like advertising, brand management, and public opinion research have greatly contributed to its meteoric popularity. Previous studies have relied mostly on classical ML-based and lexicon-based approaches, rather than deep learning (DL) methods, for classifying emotional states in English tweets. In addition, a dearth of studies analyzes the polarity of tweets written in languages other than English, such as Arabic. When dealing with massive amounts of data, as is often the case with social network data, the deep learning approach has recently exhibited outstanding performance compared to typical ML algorithms. This article aims to give readers an introduction to DL for sentiment analysis on Twitter. The TSA task is first introduced briefly. We then discuss the TSA task's framework from multiple angles, focusing on Twitter sentiment analysis. In this article, we will go through the different methods for conducting sentiment analysis and their uses and difficulties.

Keywords: *Twitter Sentiment Analysis, NLP Techniques, Opinion Mining, Social media, Deep learning, Machine learning*

1. Introduction

In today's interconnected world, social media has permeated all aspects of human interaction. Platforms like Twitter have become instrumental in sharing information and expanding knowledge among individuals, who passionately discuss various topics such as products, services, events, or places. Through multimedia messages, people succinctly express their emotions [1]. SA plays a crucial role in understanding public opinion within the realm of social media. However, this task presents significant challenges due to the immense volume of text content generated by both humans and machines [2] [3]. Twitter allows users to share their thoughts through "tweets," which can be broadcast to followers or directed to specific users. As of 2016, Twitter boasted over 313 million monthly active users, with 100 million users engaged daily [4]. Twitter sentiment analysis revolves around analyzing users' emotional states expressed in their tweets, determining the polarity of the text, and categorizing it into positive, neutral, or negative sentiments. Given the character limit of 140 imposed on tweets, users must effectively convey their messages while coping with the challenge of understanding the content and nature of specific tweets due to their sheer volume [5].

Deep learning, a subset of ML, has gained immense popularity by leveraging ANN with multiple layers to automatically learn hierarchical representations of data. This approach has proven effective in processing and extracting features from unstructured text data, including tweets. Sentiment analysis (SA), as a component of text classification, tackles the challenge of categorizing people's opinions or sentiments expressed in texts, thus enabling the identification of their interest in specific topics through the determination of positive or negative emotions [6]. In recent years, sentiment analysis has witnessed the adoption of various methods from NLP and ML. Deep learning techniques have gained prominence and achieved remarkable success in this domain. CNN and LSTM networks have emerged as particularly powerful models for sentiment analysis tasks. Extensive research has demonstrated their effectiveness, either individually or when combined. In the field of natural language processing, several methods exist for extracting meaningful features from words [7]. Word2Vec and Global Vectors for Word Representation (GloVe) stand out as two highly popular approaches. These methods provide ways to represent words as dense vectors, capturing semantic and contextual information. Word2Vec employs neural network architectures to generate word embeddings, while GloVe leverages co-occurrence statistics to create word representations. Both techniques have been widely adopted and have proven to be valuable resources for sentiment

1 School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh – 522237, India
ORCID ID : 0000-0002-2318-3327
2 School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh – 522237, India
ORCID ID : 0000-0003-0712-6899
**Corresponding Author Email: deepthi.g@vitap.ac.in*

analysis and other NLP tasks.

This paper will follow the following outline: In Part 2, we'll take a look at several related research papers. Following that, section 3 explains the significance of sentiment analysis, and section 4 delves into previously conducted research on sentiment analysis that employed conventional methods, as well as instances of previously conducted studies that utilized DL methods, along with problems and uses of TSA. The conclusion of the paper can be found in section 5.

Figures 1 and 2 show statistical analysis of the number of Twitter users per year and the population of the Twitter accounts country-wise.

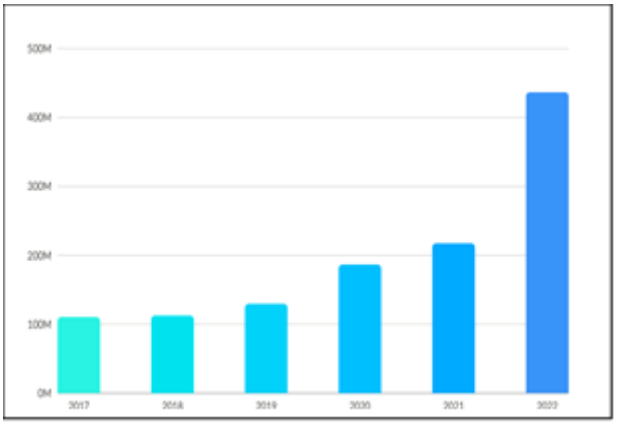


Fig 1. Twitter user growth for every year [80]

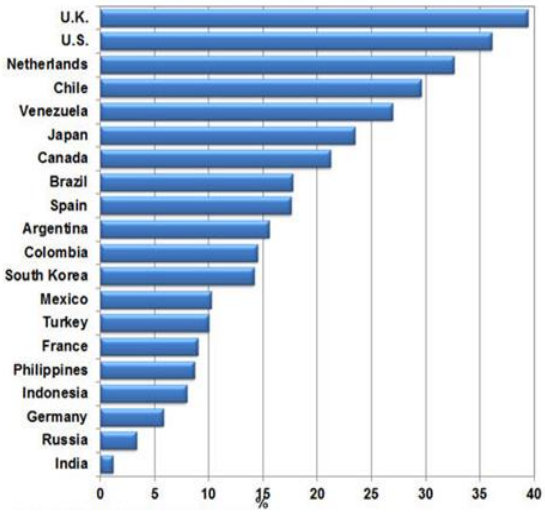


Fig. 2. The proportion of the population with Twitter Accounts [80]

2. Related Work

In their study, Sanjay et al. [8] conducted sentiment analysis on Twitter data related to the Indian farmer protests to gain insights into global public sentiment. They employed algorithms to analyze approximately twenty thousand tweets associated with the protests and assess the sentiments expressed. The researchers analyzed and contrasted the success of 2 popular text representation techniques BoW and TF-IDF, and discovered that BoW outperformed TF-IDF in sentiment analysis accuracy. The study further involved the application of various classifiers, including SVM, RF, DT, and NB, on the dataset. The results revealed that the RF classifier achieved the best possible accuracy among the evaluated classifiers.

In their research, Behl et al. [9] gathered tweets related to various natural disasters and categorized them into three groups based on their content: "resource availability," "resource requirements," and "others." To accomplish this classification task, they employed a Multi-Layer Perceptron (MLP) network with an optimizer. The proposed model demonstrated an accuracy of 83%, indicating its effectiveness in accurately classifying the tweets into the designated categories.

In their study, Tan et al. [10] introduced a model that combined BI-LSTM, RoBERTa, and GRU models. To further enhance the general effectiveness of sentiment analysis, the model's predictions were averaged using majority voting. Addressing the challenges posed by unbalanced datasets, the researchers enhanced the data by utilizing GloVe pre-trained word embeddings. The experimental results demonstrated that the proposed model surpassed state-of-the-art approaches, achieving accuracy rates of 0.942, 0.892, and 0.9177 on the Sentiment 140, USAirlines, and IMDB datasets, respectively.

For Aspect-level SA, Lu et al. [11] presented IRAN (Interactive Rule Attention Network). To simulate the operation of grammar at the sentence level, IRAN includes a grammar rule encoder that normalizes the result of adjacent locations. Furthermore, IRAN makes use of an attention network that interacts with its environment to better understand the target and its surroundings. We show that IRAN learns informative features successfully and beats baseline models by experimenting on the ACL 2014 Twitter & SemEval 2014 datasets. As a result of these results, it is clear that IRAN is an effective tool for aspect-level sentiment analysis, which can lead to enhanced performance in the field.

In their study, Mehta et al. [12] conducted a relative investigation of SA specifically focused on big data. They identified six types of emotions, namely happy, sad, joy, surprise, disgust, and fear. Additionally, they judged various methods for emotion identification that can serve as potential avenues for future research in the field. This analysis provides valuable insights into sentiment analysis in the context of big data, offering a foundation for exploring emotion identification techniques and their applications.

He et al. [13] introduced LGCF, a multilingual learning paradigm that emphasized active learning in both global and local contexts. Unlike its predecessors, this model, LGCF

demonstrated the ability to effectively learn the connections between target aspects and local contexts, along with the connections between target aspects and global contexts simultaneously. This innovative approach enables the model to capture and utilize both local and global contextual information efficiently, enhancing its overall performance in sentiment analysis tasks.

In their study [14], an extensive evaluation of sentiment polarity classification methods was specifically designed for Twitter text. Notably, they expanded the comparison by including a combination of classifiers in their analysis and introduced the aggregation and utilization of manually annotated tweets for method evaluation. This aspect is considered a significant contribution because previous attempts at automated annotation based on features like emoticons have proven problematic. Such automated approaches often fail to accurately capture the overall sentiment expressed by the author, particularly when considering instances of neutral sentiment or the absence of sentiment in the text. The inclusion of manual annotations addresses this limitation and adds value to the evaluation process of SA methods for Twitter text.

To better understand the state of the art in SA using DNNs and CNNs, Qurat et al. [15] undertook a systematic literature review of current studies. Topics covered in their investigation of sentiment analysis included text sentiment categorization, cross-lingual analysis, and both textual and visual analysis. Datasets were culled from a wide range of social media platforms. The authors presented the various stages of the successful construction of DL models in emotion analysis and noted that many difficulties in this field were efficiently solved with high accuracy using deep learning methodologies. With their more complex structures, deep learning networks were able to extract and represent features more accurately than traditional neural networks and SVMs. This study demonstrates the benefits of using DL models for sentiment analysis, which can lead to improved results in emotion analysis.

3. Sentiment Analysis

Text mining is used in the field of research known as opinion mining, which sometimes goes by the name sentiment analysis (SA). SA is a branch of study that analyzes people's perspectives and sentiments towards particular topics or events. A wide variety of activities fall under its purview, and it is referred to by a variety of names, including affect analysis, subjectivity analysis, review mining, sentiment analysis, opinion extraction, sentiment mining, opinion mining, review mining, and sentiment mining, to name a few [30].

SA involves the classification of text into positive, negative, or neutral categories. Its purpose is to analyze people's opinions in a manner that can aid business growth. Apart from polarity (positive, negative, and neutral), sentiment analysis also considers emotions such as happiness, sadness, anger, and more. To achieve this, it leverages different Natural Language Processing algorithms, including rule-based, automatic, and hybrid approaches. Fig 3 shows the SA with Polarities.

Fig 3. Sentiment Analysis

3.1. Sentiment Classification Levels

SA can be performed at various levels, including the document level, sentence level, and aspect/feature level as shown in Fig 4.

3.1.1. Document Level Classification

During this process, sentiment is extracted from the entire review, allowing for the classification of the opinion as a whole based on the overall sentiment expressed by the reviewer. The objective is to categorize the review as positive, negative, or neutral.

For instance,

"I went a Greece a couple of days ago. It was a very good place, although the cottages are so good to stay. The food was so tasty. I like that place!"

Based on the given review, the classification of the sentiment is positive. Document-level classification is most effective when the document is authored by a single individual and conveys an opinion or sentiment regarding a single entity.

3.1.2. Sentence Level Classification

This process typically consists of two steps:

- Classifying sentences into two categories: objective and subjective, known as subjectivity classification.
- Classifying subjective sentences into either positive or negative categories, referred to as sentiment classification.

Objective sentences convey factual information, while subjective sentences express personal feelings, views, emotions, or beliefs. To identify subjective sentences,

different approaches like Naïve Bayesian classification can be utilized. However, it is not enough to solely determine whether a sentence has a positive or negative opinion. This intermediate step helps filter out non-opinionated sentences and assists in assessing, to some extent, the positive or negative sentiments towards entities and their aspects. A subjective sentence can encompass multiple opinions and contain a mix of subjective and factual clauses.

For example,

"iPhone sales are flourishing despite the unfavourable economic conditions."

While SA at the document and sentence levels is valuable, it does not uncover specific preferences or dislikes of individuals, nor does it identify the targets of opinions.

3.1.3. Aspect/Feature Level Classification

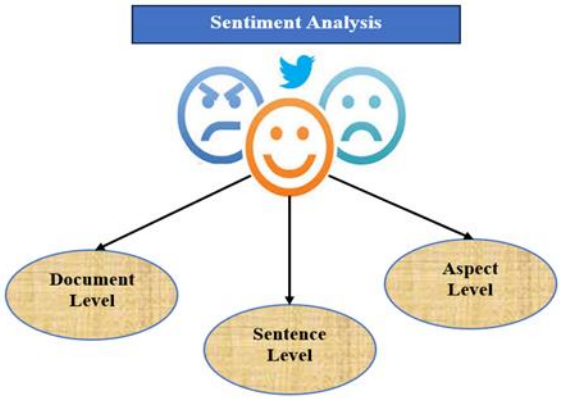


Fig 4. Levels of SA

The objective of this process is to identify and extract the attributes or features that have been mentioned by the opinion holder and ascertain whether the opinion associated with those features is positive, negative, or neutral. Synonyms of features are grouped, resulting in a feature-based summary derived from multiple reviews.

3.2. Types of SA

Sentiment analysis systems can be categorized into several different types as shown in Fig 5.

3.2.1. Fine-grained SA

Highly positive and highly negative sentiment indicators are just two of the many subcategories that can be generated by these sentiment analysis algorithms. This method is similar to using a rating scale from one to five stars to assess customer satisfaction surveys, and it is effective.

3.2.2. Emotion Detection Analysis

These sentiment analysis systems focus on identifying specific emotions rather than simply categorizing positivity and negativity. They can detect a range of emotions such as happiness, frustration, shock, anger, and sadness.

3.2.3. Intent-based Analysis

Rather than simply recognizing opinions, these sentiment analysis systems can also determine the underlying motivations behind communication. They can tell, for instance, that someone who complains online about how difficult it is to replace a battery probably wants to talk to customer support about it.

3.2.4. Aspect-based Analysis

To determine if something is being spoken of favourably or unfavourably, these sentiment analysis systems examine its constituent parts. If, in a product review, a customer complains that the battery life is too short, the sentiment analysis algorithm will correctly determine that the customer's complaint is related to the battery life and not the product as a whole.

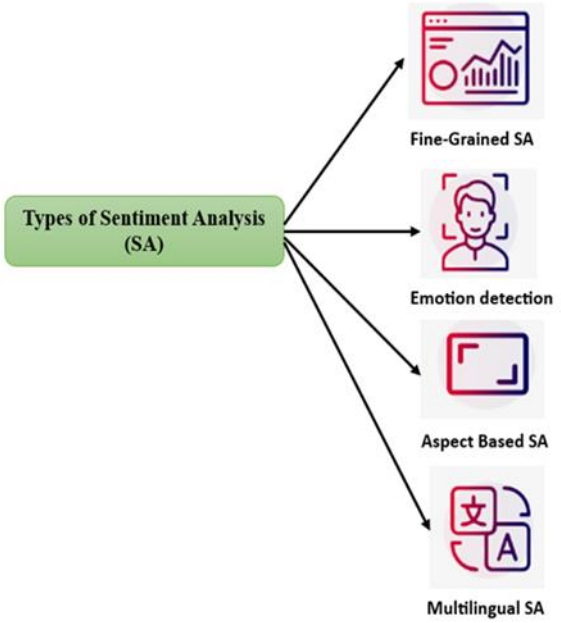


Fig. 5. Types of SA

4. Twitter Sentiment Analysis

Within the broader field of sentiment analysis, TSA holds a distinct position and serves as a prominent area of research in computational semantics. The tasks involved in sentiment analysis encompass assessing the degree of polarity indicated in a text (e.g., positive, negative, neutral), recognizing the specific target or subject of the sentiment, associating sentiment with the relevant entity or individual, and determining sentiment for different aspects of a particular topic, product, or organization [31].

The primary objective of conducting TSA is to accurately classify tweets into different sentiment categories. In this research domain, numerous techniques have emerged, providing methodologies for training and testing models to evaluate their effectiveness. However, performing sentiment analysis on Twitter tweets presents unique

challenges. In the following discussion, we outline several reasons for these challenges [32].

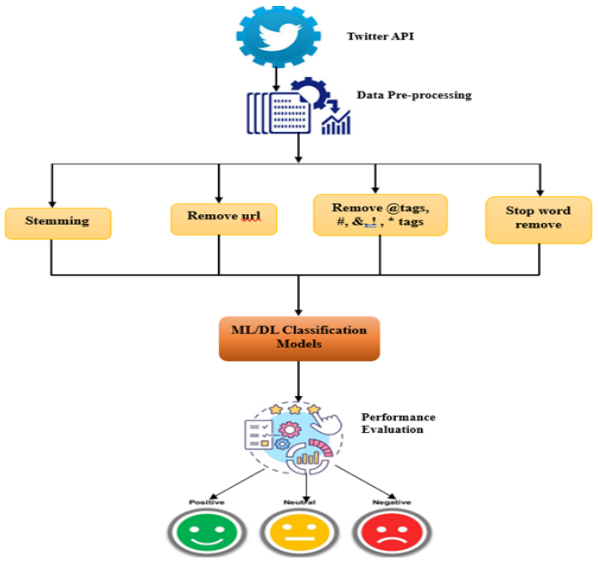
- Limited tweet size
- Use of slang
- Features of Twitter
- User Variety

The aforementioned challenges need to be addressed during the pre-processing phase.

Fig 6 shows the entire process of how we will classify Twitter tweets by using ML/DL algorithms, it follows some steps while performing TSA.

Fig 7 explains what classification models are available for sentiment analysis tasks like supervised, Unsupervised, hybrid and lexicon and also different approaches for doing TSA while performing any real-time application scenarios.

Fig. 6. Workflow of Twitter Sentiment Analysis.



4.1. Classification Models

4.1.1. Machine Learning (ML) Approaches for TSA

By gathering and cleaning data, extracting features, training data with the classifier, and analysing outcomes, ML algorithms can build classifiers to finish sentiment categorization via feature vectors [33].

Using ML techniques, the dataset must be divided into a test group and a training group. The classifier is taught on training sets to recognise specific patterns in the text, and its effectiveness is measured on a test set. Classifiers (such as the NB classifier, the SVM classifier, the Logistic classifier, and the RF classifier) divide texts into categories. Researchers frequently employ machine learning as one of the most popular approaches to text classification.

The effectiveness of sentiment classifiers on Twitter is primarily dependent on the quantity of training data as well as the feature sets that are extracted from that data. The use of SVM and NB classifiers, in particular, has become increasingly common in TSA methodologies. These strategies rely on ML approaches. The process of supervised ML algorithms for analysing sentiment on Twitter is depicted in Figure 8, which provides further explanation.

To determine the polarity of Twitter sentiment, Anton and Andrey [34] created a model. Words with n-grams and emoticons were used to extract the features. The results of the experiment showed that SVM was superior to Naive Bayes.

The SVM coupled with unigram feature extraction yielded the greatest overall performance, with an accuracy of 81% for precision and 74% for recall.

Among the four popular classifiers, namely NB, SVM, K-NN, and C4.5, the SVM classifier demonstrates superior performance across three datasets when different preprocessing methods are applied [35].

Naive Bayes algorithm for SA using data collected from Twitter.

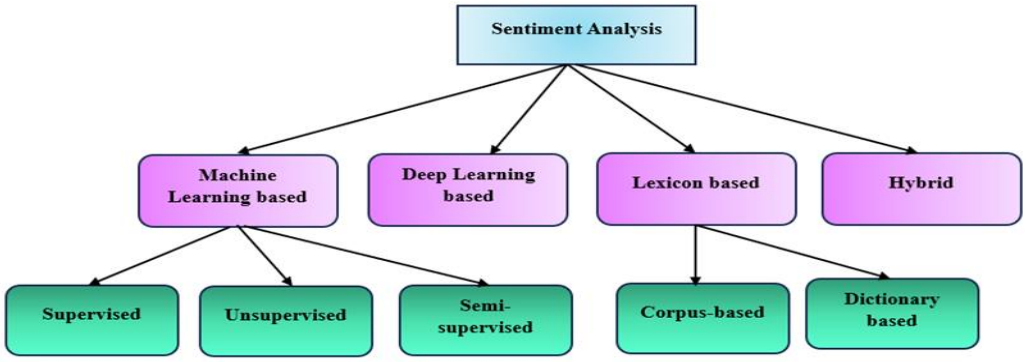


Fig. 7. SA Approaches

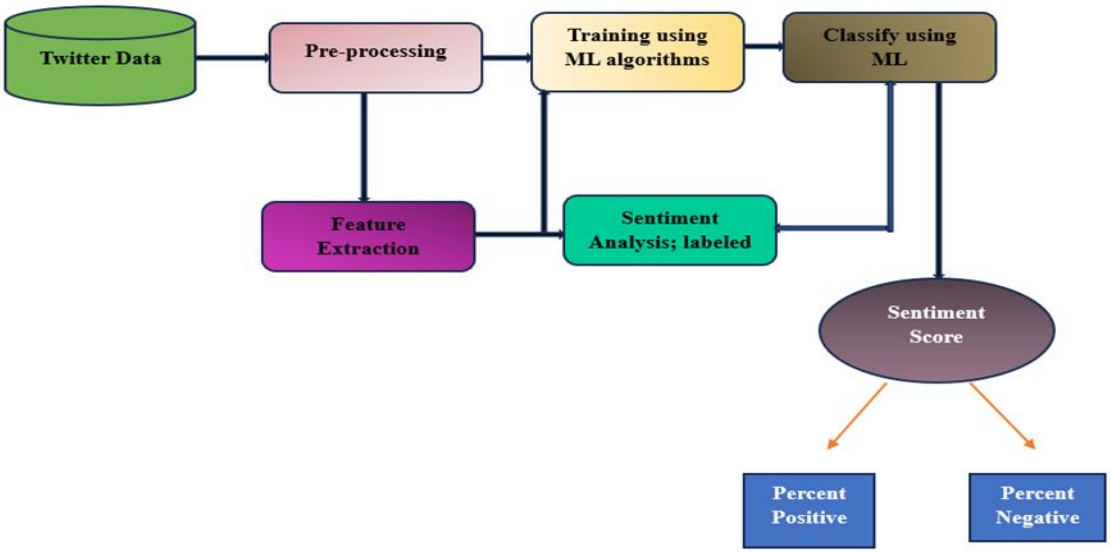


Fig. 8. Process of TSA using ML models

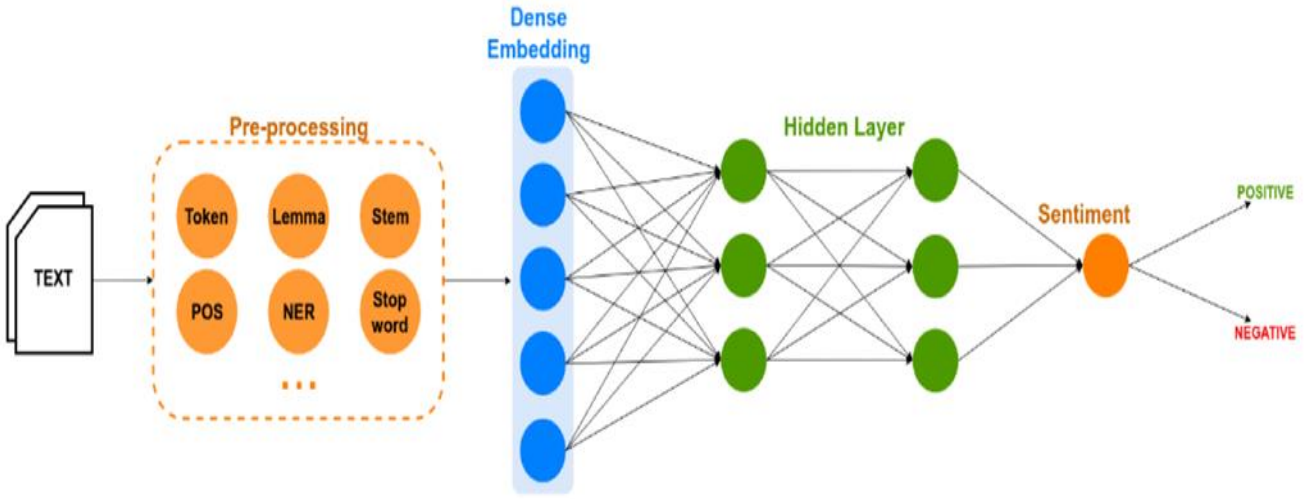


Fig. 9. SA process in DL [81]

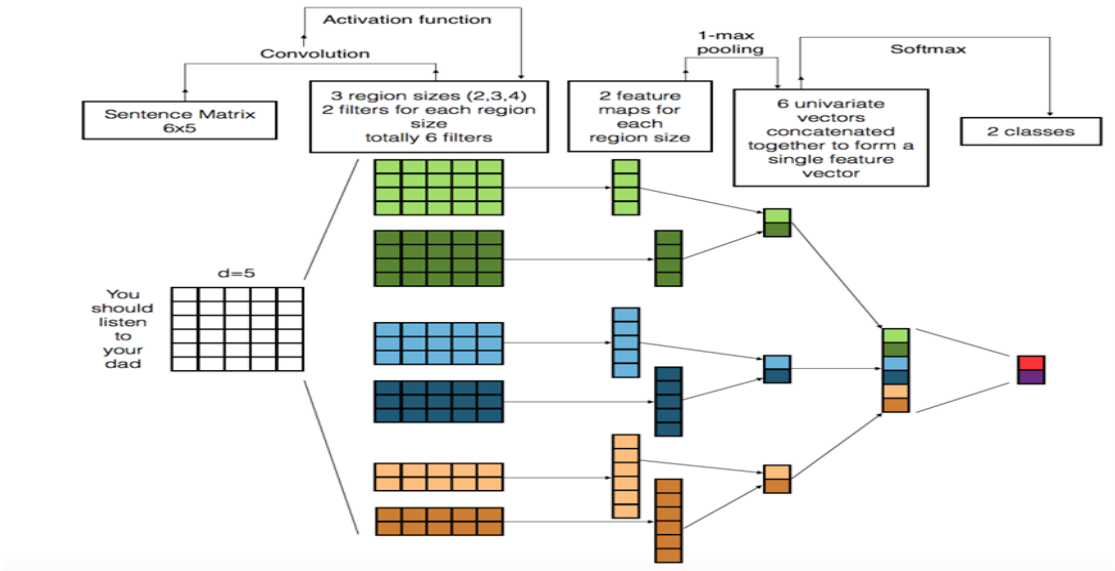


Fig. 10. TSA using CNN [82]

The research involves a data crawler that retrieves Twitter data for analysis. The Naive Bayes algorithm is employed to classify the sentiment of the Twitter data into positive, negative, or neutral categories. The study conducted a comparison of the NB, SVM, and K-NN methods using RapidMiner. The NB method achieved an accuracy of 75.58%, while the SVM method demonstrated an accuracy of 63.99%, and the K-NN method achieved an accuracy of 73.34% [36].

4.1.2. Deep Learning (DL) Approaches for TSA

DL, a subset of machine learning, focuses on algorithms that learn multiple layers of data representations or features, leading to cutting-edge classification outcomes. The idea behind neural networks is inspired by the intricate network of thousands of neurons in the human brain, as depicted in Figure 9. Recently, deep learning has garnered significant attention and has demonstrated remarkable achievements in various domains, including NLP, computer vision, speech recognition, and more. Its impressive performance in tasks such as language comprehension and image analysis suggests that it holds the potential for achieving similar levels of accuracy and efficiency in sentiment analysis, a crucial component of language understanding [37].

A. Convolutional Neural Networks (CNN)

Unlike other types of neural networks, CNNs use a series of convolutional layers, pooling layers, dropout layers, and fully connected layers to achieve their desired results. After the input data has been processed by the convolutional and pooling layers, it is passed on to the fully connected layers for classification, where the most relevant feature representations have been extracted. Features to be recovered from the data are determined by the size and number of convolutional filters used.

These filters are applied to the full input data set, allowing for the capture of duplicate features across several coordinate systems. Then, a max-pooling layer is used to pick out the most important features from the CNN's outputs. The cleansed data is then used as input in an RNN model for further classification. In the fully linked layer, where the output is produced, the activation function softmax is typically utilized. Several filters, kernel size, padding, strides, and activation functions are some of CNN's other important factors. In the provided experiment, CNN is set up using 24 filters, a kernel size of 4, 'valid' padding, 2x2 strides, and the ReLU activation function. Figure 10 depicts the full procedure of a CNN.

B. Long short-term memory networks (LSTM)

LSTM is a specialized type of RNN that excels at processing sequential data like time series, speech, and text. With its ability to capture long-term dependencies, LSTM networks are particularly effective for tasks such as language translation, speech recognition, and time series forecasting.

LSTM neural network is designed in Fig 11 to handle sequential data [38], while effectively mitigating the vanishing error gradient problem. Additionally, it excels at capturing long-term dependencies through its gated structure. Mathematically, we can represent it as follows:

$$h_{t\,\mathrm{h}} = f(W_h . x_t + U_t . h_{t-1} + b_h) \quad (1)$$

$$f_{t\mathrm{h}} = \sigma(W_f . x_t + U_f . h_{t-1} + b_f \quad (2)$$

$$i_{t\mathrm{h}} = \sigma(W_i . x_t + U_i . h_{t-1} + b_i) \quad (3)$$

$$o_{t\mathrm{h}} = \sigma(W_o . x_t + U_o . h_{t-1} + b_o) \quad (4)$$

$$c_{t\mathrm{h}} = f_t \circ c_{t-1} + i_t \circ \tanh(W_c . x_t + U_c . h_{t-1} + b_c) \quad (5)$$

$$h_{t\mathrm{h}} = o_t \circ \tanh(c_t) \quad (6)$$

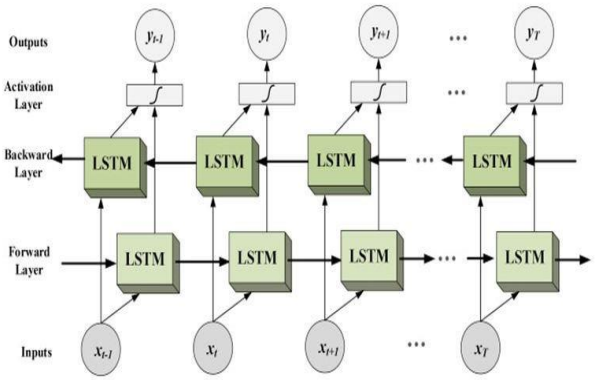


Fig. 11. LSTM Process for TSA

The aforementioned equations are associated with LSTM network mathematical calculations.

Bidirectional LSTM (BiLSTM) is an enhanced version of the LSTM architecture. It comprises two LSTM units, with one unit processing the input sentence in a forward direction, and the other unit reading it in a backward manner. The hidden states of each LSTM unit are then concatenated to form the final word representation [39].

$$h_t^{\,bilstm} = h_t^{\,forward} \oplus h_t^{\,backward} \quad (7)$$

where $\oplus$ denote the operator of the concatenation.

C. Recurrent Neural Network (RNN)

RNNs, possess the ability to retain information from previous inputs by utilizing an internal memory, making them suitable for handling sequential data like text, genomes, or numerical time series data [40, 41].

Predicting the next word in a sequence by considering the preceding words is a challenging task for traditional neural networks. However, RNNs overcome this challenge by

employing hidden layers that retain sequential information over time, enabling them to generate output [42, 43].

Every hidden layer in a recurrent neural network (RNN) possesses its weight and is accompanied by a sigmoid activation function. However, RNNs are limited to addressing short-term dependencies and struggle with handling long dependencies due to the vanishing gradient problem that arises during backpropagation through time [44].

D. Other Neural Networks

Deep belief networks (DBNs) [45] are a type of deep neural network characterized by multiple layers of a graphical model that includes both directed and undirected edges. Each DBN consists of interconnected layers of hidden units, with connections between consecutive layers but no connections within a layer. Learning a DBN involves employing a greedy layer-wise learning algorithm.

Recursive neural networks (RecNNs) [46] can be seen as an extension of RNNs and are commonly used to learn directed acyclic graph structures from data. In a RecNN, the hidden state vectors of the graph's left and right child nodes are utilized to compute the hidden state vector of the current node.

Hybrid deep learning [47] represents another category that involves combining multiple deep learning techniques. Examples include the fusion of CNNs with LSTM [48], or the combination of probabilistic neural networks (PNNs) with a two-layered restricted Boltzmann machine (RBM).

The ML and DL in Twitter sentiment analysis rely on a pre-existing list of positive and negative words to determine the polarity of a message or the sentiment expressed by the individual being analyzed. Table 2 presents examples of Twitter sentiment analysis studies that have utilized ML/DL methods.

4.1.3. Ensemble Approaches for TSA

Ensemble methods, which involve the combination of multiple classifiers, are employed to enhance the accuracy and precision of predictions. In the realm of text classification, particularly Twitter sentiment analysis, leveraging ensemble methods can offer significant benefits by improving the classification accuracy of tweets.

The main challenge in Twitter sentiment analysis lies in finding the optimal sentiment classifier that can accurately categorize tweets. Typically, popular base classifiers such as NB, RF, SVMs, and LR are employed, introducing an ensemble classifier that combines these base classifiers into a unified classifier, aiming to enhance the performance and accuracy of SA [53].

Recent research has shown that sentiment analysis of tweets can be useful in a wide range of settings. For instance, in work [54], the authors analyzed sentiments expressed on Twitter about getting vaccinated against pneumonia viruses right away between December 2021 and July 2021. Several DL techniques, such as RNN, LSTM, and bidirectional LSTM, were compared and contrasted for their efficacy. The highest levels of accuracy were obtained with LSTM (90.59%) and Bi-LSTM (90.83%). Aspect-based sentiment analysis was used with six Twitter emotions and four different BERT models [57] in a different study. The maximum accuracy of 87% was obtained using the proposed strategy. In [58], 54,065 tweets in Arabic were processed using four classifiers RF, gradient boosting (GB), K-NN, and SVM. An ensemble method was used to combine the four classifiers, resulting in an accuracy of 89.12%. Finally, [59] analyzed tweets about the COVID-19 vaccination in the US to see how vaccine resistance develops over time.

Twitter sentiment analysis employs an ensemble-based approach that utilizes a pre-existing list of positive and negative words to ascertain the polarity of a message or the sentiment expressed by the individual under analysis. Table 3 provides examples of studies in Twitter sentiment analysis that have adopted ensemble-based methods.

4.1.4. Lexicon-based Approaches for TSA

Following these guidelines is the basis of the lexicon-based approach: Three steps are involved in determining the polarity of a piece of text: (1) breaking down sentences into words, (2) matching those words to entries in a sentiment polarity lexicon, and (3) generating an overall score. Through this method, we may determine whether a piece of writing is positive, negative, or neutral in tone [66]. Words in the lexicon-based approach have their semantic orientation determined through the use of a dictionary or a corpus. The former is more straightforward and requires only the use of a sentiment dictionary containing opinion terms to ascertain the polarity score of words and phrases in the text.

In their study, Asghar et al. [67] introduced an enhanced lexicon-based sentiment classification approach that integrated a rule-based classifier. The aim was to address data sparsity issues and enhance sentiment classification accuracy. The approach sequentially incorporated various classifiers, including those utilizing emoticons, modifier-negation, Sentiment WordNet (SWN), and domain-specific classifiers. This sequential integration enabled accurate sentiment classification of tweets based on their polarity. The invented model got impressive F1 scores of 0.8%, 0.795%, and 0.855% for three distinct datasets comprising drug, car, and hotel reviews, respectively.

The lexicon-based approach in TSA relies on a set of positive and negative words that can be used to identify the

tone of a communication or the sentiment expressed by the individual being analyzed. Table 4 presents examples of Twitter sentiment analysis studies that have utilized lexicon-based methods.

4.2. Datasets

Training ML algorithms can greatly benefit from access to social media data. Countless datasets may be accessed through open sources such as Kaggle, the UCI repository, and others. The exclusion of such information was deliberate, as it was unnecessary for the testing process. Below, we provide a few dataset descriptions, which are very popular for doing TSA.

- The Sentiment140 dataset, sourced from Stanford University [48], comprises 1.6 M tweets discussing products or brands. The dataset was pre-labelled, associating each tweet with a polarity indicating the sentiment expressed by the author (0 for negative sentiment, 4 for positive sentiment).
- The Tweets Airline dataset [49] is a collection of tweets expressing user opinions specifically related to U.S. airlines. This dataset was obtained through web crawling in February 2015. It consists of 14,640 samples, which have been categorized into negative, neutral, and positive classes.
- The IMDB Movie Reviews dataset [50] comprises comments from audiences discussing the narratives of films. The dataset consists of 25,000 samples, which have been categorized into positive and negative sentiments.

4.3. Pre-processing

Once the textual data from Twitter is collected, the subsequent step is pre-processing, which is implemented in Python. The pre-processing stage involves multiple steps, starting with the conversion of uppercase letters to lowercase. This step ensures uniformity in the text data and eliminates any potential discrepancies arising from the varying capitalization of letters [51].

For eg: FACEBOOK to facebook

Filtering out irrelevant, inaccurate, or otherwise undesirable information is the essence of data pre-processing. It includes the following steps:

- In the case of a Twitter dataset, this includes: removing retweets;
- Stripping out URLs, symbols, punctuation, numbers, etc.;
- Stemming and tokenizing the text; and
- Stopword removal

4.4. Feature Extraction

4.4.1. Bag-of-Words (BoW):

BoW represents text as a collection of unique words, disregarding their order and focusing on their frequencies. Each document or tweet is represented by a vector where each element corresponds to the occurrence or frequency of a specific word. Stop words and punctuation marks are often removed to reduce noise.

Mathematically, the Bag-of-Words representation for a document can be represented as

$BoW(d) = [w1, w2, w3, ..., wn]$

Where:

BoW(d) represents the Bag-of-Words representation for document d.

w1, w2, w3, ..., wn represent the frequency or presence of each word in the vocabulary.

For example, consider the following sentence: "The cat chased the mouse."

Using BoW representation, we tokenize the sentence and create a vocabulary: ["The", "cat", "chased", "mouse"].

Then, we count the frequency of each word in the sentence: [1, 1, 1, 1].

Finally, we represent the sentence as a vector using the word frequencies: [1, 1, 1, 1].

Note that in some variations of BoW, instead of frequencies, binary values (0 or 1) may be used to indicate the presence or absence of a word in the document.

4.4.2. TF-IDF:

TF-IDF computes a word's significance in a document by factoring in its frequency within the document (term frequency) and its scarcity across the entire collection of documents (inverse document frequency). This technique assigns higher weights to words that are more discriminative for sentiment classification.

TF-IDF scores are calculated using the following step-by-step formulas.

$$tf(w,d)=\log(1+fw,d) \quad (8)$$

$$idf(w,D)=\log(N/f(w,D)) \quad (9)$$

$$tfidf(w,d,D)=tfw,d*idf(w,D) \quad (10)$$

4.4.3. Word Embeddings:

Captures the semantic meaning of words by representing them as dense vector representations in a continuous space. Popular algorithms like Word2Vec, GloVe, and FastText learn these embeddings from large text corpora. In sentiment analysis, word embeddings can capture contextual information and relationships between words.

4.4.4. N-grams:

Sequences of N words that span the same section of text. N-grams are useful because they take into account word sequences rather than single words, thus capturing contextual information and dependencies. Single-word models (unigrams), double-word models (bigrams), and triple-word models (trigrams) all fall under the general category of N-gram models.

Example: “I stay in Madras”

Unigram [“I”, “stay”, “in”, “Madras”]

Bigram [“I stay”,”stay” in” “in Madras”]

Trigram [“I stay in”, “stay in Madras”]

Table 1. Different Implemented datasets used in Existing work for Twitter SA

S.No	Dataset Name	Description	Source
1	Sentiment Self-driving Cars	In a straightforward task of Twitter sentiment analysis, participants were assigned the job of reading tweets and categorizing them as very positive, slightly positive, neutral, slightly negative, or very negative. Additionally, they were prompted to indicate if the tweet was unrelated to self-driving cars.	https://www.crowdflower.com/data-for-everyone/

2	Covid-19 pandemic for Binary sentiment analysis	The objective of this dataset is to facilitate sentiment analysis on tweets related to the COVID-19 pandemic. The dataset consists of three versions, containing 186,000, 132,000, and 82,000 English tweets respectively, with stopwords removed. In all versions, tweets with a polarity of 1 are considered positive, while those with a polarity of 0 are deemed negative.	https://data.mendeley.com/datasets/t8bxg423yk/1

3	Entity-level Twitter Sentiment Analysis	The provided dataset is designed for entity-level Twitter sentiment analysis, where the goal is to determine the sentiment of the entire sentence towards a specific entity. For instance, a sentence like "A outperforms B" would be considered positive for entity A but negative for entity B. The dataset comprises approximately 70,000 labeled training messages and 1,000 labeled validation messages. It can be freely accessed on Kaggle's online platform.	-
4	Arabic Sentiment Analysis	Using a tweet crawler, they gathered 2K labeled tweets (1K positive and 1K negative) on diverse subjects like politics and arts.	https://archive.ics.uci.edu/ml/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis

5	Entity-level Sentiment Analysis on Multi-Lingual tweets	The provided dataset focuses on entity-level sentiment analysis of Twitter data. The purpose of this task is to evaluate the tone of a given message about a particular entity. The dataset consists of three classes: Positive, Negative, and Neutral. Messages that are deemed irrelevant to the entity are categorized as Neutral.	Kaggle.com

6	BTC Tweets Sentiment	The dataset comprises Bitcoin-related tweets extracted from a daily sample, which were initially assigned sentiment scores. The last two columns, namely "New_sentiment_score" and "New_sentiment_state," were added based on predictions made by a trained NLP model. These additional columns serve as a comparison to the original sentiments.	data.world
7	Sentiment 140	Sentiment140 enables users to explore the sentiment surrounding a brand, product, or topic on Twitter.	Tensorflow datasets

8	AfriSenti-Twitter-Sentiment	AfriSenti stands as the largest benchmark dataset for sentiment analysis in under-represented African languages. It encompasses over 110,000 annotated tweets across 14 African languages.	Huggingface.com
9	Twitter US Airline sentiment	The task involved conducting sentiment analysis on the issues faced by major U.S. airlines. Twitter data from February 2015 was collected, and contributors were instructed to classify tweets as positive, negative, or neutral. Furthermore, they were asked to categorize the reasons behind negative sentiments, such as "late flight" or "rude service." The unaggregated results, consisting of 55,000 rows, are available for download.	-

4.4.5. Part-of-Speech (POS) Tags:

POS tagging involves assigning grammatical tags (such as noun-verb-adjective) to each word in a sentence. These tags provide information about the syntactic structure of the text, which can be useful for sentiment analysis. For example, adjectives and adverbs often carry sentiment information.

4.4.6. Sentiment Lexicons:

Contain pre-defined sentiment polarity scores for words. These lexicons can be used to assign sentiment scores to individual words or calculate overall sentiment scores for documents or tweets. Examples of sentiment lexicons include AFINN, SentiWordNet, and VADER.

4.4.7. Deep Learning-Based Features:

With the rise of DL, features can be automatically learned from raw text data using neural network architectures. Techniques such as CNNs and RNNs can capture hierarchical and sequential patterns in text, extracting informative features for SA.

	airline_sentiment	text
0	neutral	@VirginAmerica What @dhepburn said.
1	positive	@VirginAmerica plus you've added commercials t...
2	neutral	@VirginAmerica I didn't today... Must mean I n...
3	negative	@VirginAmerica it's really aggressive to blast...
4	negative	@VirginAmerica and it's a really big bad thing...

	airline_sentiment	text
0	neutral	@VirginAmerica What @dhepburn said.
1	positive	@VirginAmerica plus you've added commercials t...
2	neutral	@VirginAmerica I didn't today... Must mean I n...
3	negative	@VirginAmerica it's really aggressive to blast...
4	negative	@VirginAmerica and it's a really big bad thing...

Fig. 12. Sample tweets from the dataset

Table 2. ML and DL Approaches for TSA

Ref	Model	Datasets	Accuracy in(%)	Limitation
[16]	NB, SVM and DT	Publicly available	The overall accuracy of 91	Assumes independence between features
[17]	BPNN, SVM	Kaggle dataset	69.24.00 66.85	Requires careful tuning of hyperparameters
[18]	NB	Self Constructed	0.83	
[19]	SVM, LibLinear Model	Self Constructed	0.98	Low efficiency when there is more background noise in the data, as when the target classes overlap.
[20]	SGD, SVM, NB, DT, RNN, K-NN	SemEval	85% with CNN and MAF	Due to issues with vanishing gradients, simple RNNs struggle to process lengthy sentences.

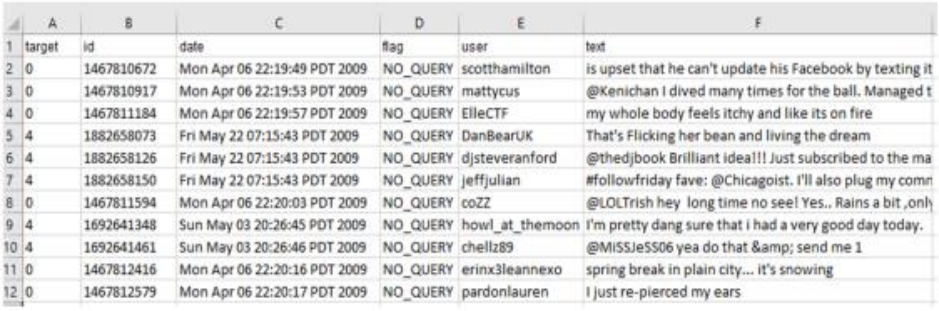
[21]	NB, SVM, Maximum Entropy	IMDB, Amazon, Airlines twitter dataset	0.88, 0.87, 0.93 (LSTM)	NB does not consider the context or order of words in a sentence, which can result in a loss of important information for sentiment analysis. It treats each feature as an independent entity, disregarding the sequential nature of text data
[22]	NB, SVM	Self Constructed	86.5%,72.6% 87.96%	SVMs can become computationally expensive, especially when working with large datasets or high-dimensional feature spaces. Training an SVM model can be time-consuming, particularly with non-linear kernels.
[23]	SVM	Self Constructed	0.75%	Difficulty Handling Imbalanced Datasets

[24]	NB, SVM, WordNet, Maximum Entropy	Self Constructed	0.74%	NB assumes that the features (words or tokens) are independent of each other given the class label. This assumption may not hold true in real-world scenarios, as words in a sentence can have complex relationships and dependencies
[25]	SVM, NB	Twitter Dataset	WordNet:0.89%	SVM is computationally expensive, especially when working with large datasets

[26]	LSTM, SPARK	Sentiment140, Twitter Dataset	Positive: 83.12% Negative: 78.8%	LSTMs are designed to handle sequential data by capturing long-term dependencies, they may still struggle with capturing very long-range dependencies. Extremely long sentences or documents may pose challenges for LSTMs in effectively capturing relevant contextual information.

[27]	SVM, CNN	STS-Test, STS-Gold, SS-Twitter SE-Twitter	86.53%	CNNs typically require fixed-length inputs, which can be challenging when dealing with variable-length texts such as tweets or user-generated content. Preprocessing techniques like padding or truncation may be required, potentially leading to information loss or mismatched contexts.
[28]	CNN	Twitter Dataset, Product Data review	75.16% 65.96%	CNNs rely on pre-trained word embeddings, which may not adequately handle out-of-vocabulary words or rare terms

[29]	LSTM, DCNN, NB, SVM	Twitter Dataset	LSTM: 76.40% DCNN: 76.44%	Deep CNNs rely heavily on the quality and representation of input features. The choice of word embeddings or other input representations can significantly impact the model's performance. Inadequate or suboptimal input representations may result in a loss of important sentiment-related information.



	A	B	C	D	E	F
1	target	id	date	flag	user	text
2	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by texting it
3	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Managed t
4	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
5	4	1882658073	Fri May 22 07:15:43 PDT 2009	NO_QUERY	DanBearUK	That's Flicking her bean and living the dream
6	4	1882658126	Fri May 22 07:15:43 PDT 2009	NO_QUERY	djsteveranford	@thedjbook Brilliant idea!!! Just subscribed to the ma
7	4	1882658150	Fri May 22 07:15:43 PDT 2009	NO_QUERY	jeffjulian	#followfriday fave: @Chicagoist. I'll also plug my comn
8	0	1467811594	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	coZZ	@LOLTrish hey long time no see! Yes.. Rains a bit ,onl
9	4	1692641348	Sun May 03 20:26:45 PDT 2009	NO_QUERY	howl_at_themoon	I'm pretty dang sure that i had a very good day today.
10	4	1692641461	Sun May 03 20:26:46 PDT 2009	NO_QUERY	chellz89	@MiSSJeSS06 yea do that & send me 1
11	0	1467812416	Mon Apr 06 22:20:16 PDT 2009	NO_QUERY	erinx3leannexo	spring break in plain city... it's snowing
12	0	1467812579	Mon Apr 06 22:20:17 PDT 2009	NO_QUERY	pardonlauren	I just re-pierced my ears

Fig. 13. Sample example for Sentiment 140 dataset

These are just a few examples of feature extraction techniques used in SA.

Figure 12 and Figure 13 display sample tweets from various datasets. The tweets in these datasets are visualized using word clouds, where positive and negative tweets are depicted separately. Figure 14 highlights the wc d representation of positive and negative tweets (14) e datasets.

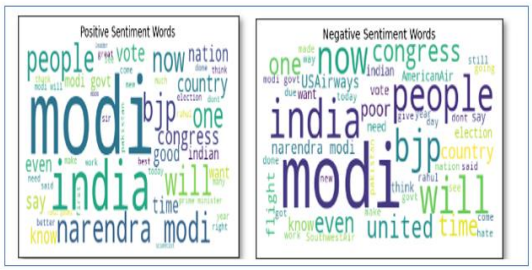


Fig. 14. Sample Word cloud from the dataset

Sample example for Predicted Sentiments from the Tweets

1. *"I dislike having to make phone calls and interrupt people's sleep"*

In the above example, the predicted sentiment is negative.

2. *"The food was meh"*

In the above statement, the predicted sentiment is neutral.

3. *"He is the greatest minister India has ever had"*

In the above tweet, the predicted sentiment is positive.

4.5. Evaluation Metrics

Our objective is not focused on creating a predictive model, but rather on the validation and selection of the best model using out-of-sample data. It is crucial to thoroughly validate the model's performance before calculating predictive values. The evaluation metrics play a significant role in quantifying the performance of the predictive model. When building the model, selecting the appropriate statistical metric is essential as it influences the evaluation and comparison of machine learning algorithm performance [52]. Moreover, the choice of metrics impacts the consideration of various characteristics in the results and ultimately affects the decision-making process for selecting the most suitable algorithm. In the context of classification problems, there are numerous statistical indicators available for investigation and analysis. Fig 15 shows the Confusion matrix to evaluate the performance of the ML/DL model.

The evaluation metric of TSA is precision (P), recall (R), F1-score (F1) and accuracy (Acc).

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{13}$$

$$Acc = \frac{TP + TN}{TP + FN + TN + FP}$$

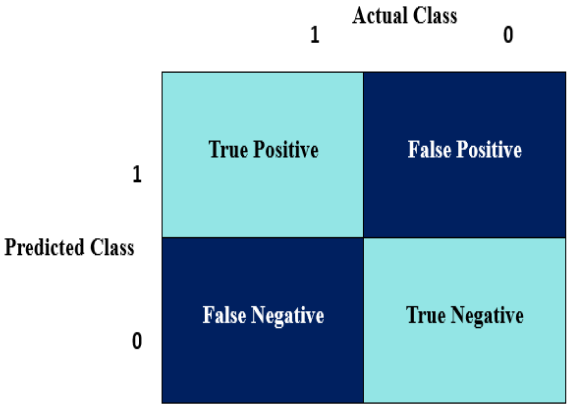


Fig. 15. Confusion Matrix

4.6. Challenges

While sentiment analysis shows promise, it has a number of obstacles that must be overcome before it can reach its full potential. One difficulty is that, despite the progress made by automated systems, human judgment is still superior in determining emotional tone. Automatic systems sometimes fail to correctly examine the specific contextual meaning of a word, and they have trouble telling the difference between sarcastic and serious writing. Acronyms and abbreviations can also be difficult to decipher. Furthermore, it can be difficult to categorize contradictory viewpoints like "I like the food quality, but the dish I ordered was not good" and to find genuine feedback within broad statements like "I astonished my brother on his birthday with a new branded watch, and he just distracts!" An automated method is also not likely to spot biased or false product or service reviews.

Problems in analyzing emotions:

Problems that frequently arise when performing sentiment analysis include:

4.6.1. Neutral sentiments:

Neutral sentiment comments often present challenges for sentiment analysis systems, as they are prone to misidentification. For instance, consider a customer who received the wrong colour item and left a comment stating, "The product was blue." Such a comment could be mistakenly labelled as neutral when in reality it should be classified as negative.

4.6.2. Unclear language:

Identifying sentiment becomes challenging when systems lack contextual understanding and fail to grasp the intended

tone. Poll or survey responses like "nothing" or "everything" pose categorization difficulties when the surrounding context is absent; their sentiment could be labelled as positive or negative based on the specific question. This phenomenon is known as lexical ambiguity. Furthermore, training systems to accurately detect irony and sarcasm proves challenging, often leading to inaccurately labeled sentiments. Pronoun resolution presents another hurdle for algorithms, as determining the antecedent of a pronoun within a sentence can be problematic. For instance, in analyzing the comment "We went to a park and then dinner. I didn't enjoy it," the system may struggle to identify whether the writer didn't enjoy the park or the dinner.

4.6.3. Unclassifiable language:

Computer programs face challenges in comprehending emojis and distinguishing irrelevant information. To address this, careful consideration is required when training models to handle emojis and neutral data. This is crucial to prevent the improper flagging of texts by the models.

4.6.4. Ambiguous sentiments:

Individuals often express contradictory statements, leading to reviews that contain both positive and negative comments. To handle this scenario, a common approach is to analyze sentences independently. However, sentiment analysis tools can face challenges when encountering sentences that include contrastive conjunctions, which feature two contradictory words. For instance, consider the statement, "The packaging was terrible but the product was great." This type of sentence can potentially confuse sentiment analysis tools.

4.6.5. Small data sets:

SA tools exhibit optimal performance when applied to substantial volumes of text data. Smaller datasets often lack the necessary depth to provide meaningful insights.

4.6.6. Language evolution:

Language, particularly on the internet, undergoes constant changes, with users frequently employing new abbreviations, and acronyms, and exhibiting poor grammar and spelling. The ever-evolving nature and wide variation in language usage pose significant challenges for algorithms.

4.6.7. Fake reviews:

Distinguishing between genuine and fake product reviews, as well as other text generated by bots, can present challenges for algorithms.

4.6.8. Human intervention:

To ensure consistency and accuracy, even the most cutting-edge AI-driven solutions for sentiment analysis and social media monitoring require human participation, as reported by Gartner.

4.6.9. Dealing with analogies:

The bag-of-words approach lacks effectiveness in making accurate comparisons. Due to its disregard for contextual relationships, a tweet such as "IITs are better than most of the private colleges" would be interpreted as positive for both IITs and private colleges within the model, overlooking the intended comparative aspect.

4.7. Applications of SA

Sentiment analysis has wide-ranging applications across various domains, including business and marketing, politics, healthcare, and public action. It extends beyond a single application and offers a multitude of users in different fields, aiding decision-making processes.

A company's reputation is built on more than just the quality of its products. Customer service, content marketing, internet marketing, and social media initiatives are all crucial to its success. Using sentiment analysis to learn how people feel about a product, brand, or company from every viewpoint is a powerful tool. The public's perception of the company is heavily influenced by this all-encompassing perspective [72].

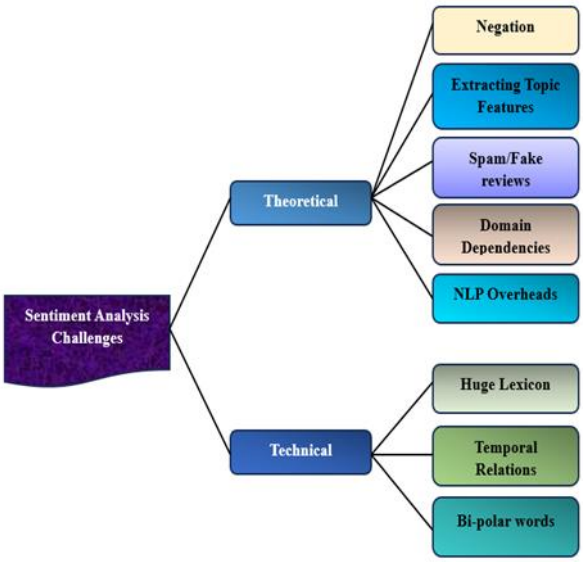


Fig. 16. Challenges of SA

The application of SA in healthcare is evident, as demonstrated by studies proposing a service framework that utilizes sentiment analysis and spatiotemporal properties to identify disease outbreak locations [73]. Furthermore, sentiment analysis can aid in identifying the emotional needs of individuals during a disaster, enabling the implementation of suitable response and rescue efforts [74]. Additionally, sentiment analysis enables the assessment of an individual's level of depression by observing and analyzing emotions expressed in text [75].

Sentiment analysis has the potential to be utilized for predicting political elections, with Twitter data proving to be a reliable platform. A notable study found a 94%

correlation between sentiment analysis results from Twitter and polling data, suggesting that it has the capability to compete with sophisticated polling techniques [76].

The feedback of customers plays a pivotal role in the application of sentiment analysis, as it enables businesses and organizations to take appropriate actions for enhancing their products, services, and overall business strategies. This significance is evident in a study that examines the opinions and experiences of social media users regarding drug and cosmetic products, providing valuable insights [77]. Additionally, sentiment analysis proves valuable in identifying areas that require improvement in airport service quality, allowing for the implementation of corrective measures, such as paying attention to passenger feedback on social media platforms [78]. Moreover, sentiment analysis facilitates the analysis of trends and characteristics in people's food habits, offering valuable insights for businesses when devising product and marketing strategies [79].

5. Conclusion

In this research, we give a comprehensive review of the ML/DL models and relevant approaches used for sentiment analysis of social network data. Several methods for assessing the emotional content of tweets are explored here. ML, ensemble methods, and lexicon-based approaches are all examples of such tactics.

While ensemble models offer several advantages for Twitter sentiment analysis, they also have some limitations to consider:

Table 3. Ensemble Approaches for TSA**Ref**	**Model**	**Datasets**	**Accuracy in(%)**	**Limitation**
[53]	Naïve Bayes Random forest SVM LR Ensemble	Twitter Sentiment Corpus	NB-75.91 RF-70.67 SVM-74.61 LR-74.15 78.19	Require additional resources in terms of computational power, memory, and training time.
[53]	Naïve Bayes Random forest SVM LR Ensemble	Healthcare	NB-71..81 RF-71.43 SVM-70.30 LR-68.92 74.59	higher risk of overfitting, particularly if the base classifiers are highly correlated or if the ensemble is trained on a limited and biased dataset

[53]	Naïve Bayes Random forest SVM LR Ensemble	First GOP Debate dataset	NB-82.20 RF-82.57 SVM-83.44 LR-81.51 83.78	The performance of ensemble models heavily relies on the diversity and quality of the base classifiers
[53]	Naïve Bayes Random forest SVM LR Ensemble	Twitter Sentiment Analysis	NB-73.65 RF-70.61 SVM-74.36 LR-74.33 80.16	increased complexity when especially dealing with large datasets.
[60]	LSTM Three-CNN Four-CNN		85.16 84.11 86.62	Integrating LSTM models into ensemble frameworks can be challenging due to the sequential nature of LSTMs and their distinct input requirements. LSTMs are prone to overfitting, especially when the training dataset is small or when the model is overly complex

	Four-CNN Features+SVM Ensemble model		85.62 85.71	
[61]	Ensemble of SVM and RF	Self-constructed	86.4	SVM and RF may be affected by noisy or imbalanced data, which can result in suboptimal performance.
[62]	SentiBank, SentiStrength	Twitter	91.32	-
[63]	SentiWordNet	Publicly available	73.64	-
[64]	ML, Lexicon	Self-constructed	86.4	Lexicon-based approaches may face difficulties in accurately identifying and handling neutral sentiment, as lexicons are primarily designed to capture positive and negative sentiment polarities.

[65]	SVM NB LR	STS and HCR	93.94 by SVM for STS and 85.09 by NB for HCR Datasets	SVM and LR perform well when the relationship between input features and sentiment labels is approximately linear. However, sentiment analysis often involves capturing more complex non-linear patterns in text data. This limitation may result in lower performance for sentiment analysis tasks that require capturing nuanced sentiment expressions.

Table 4. Lexicon-based Approaches for TSA

Ref	Model	Datasets	Accuracy in(%)	Limitation
[67]	Lexicon & dictionaries	Twitter review datasets(3)	F1-score: 79.5 for the second dataset F1-score: 85.5 for the third dataset	Difficulty with Mixed Sentiments or Emotions
[68]	Rule based classifier	Own (Self-constructed)	92 percent for binary and 87 percent for multi-class classification	Rule-based classifiers are often designed based on specific domains or datasets. They may not perform optimally when applied to different domains or contexts, as the rules may not capture domain-specific sentiment patterns effectively.
[69]	Lexicon based models	Stanford Twitter Sentiment, Obama McCain Debat	STS: 96.11 OMD: 88.84	Lexicon-based approaches may struggle to accurately capture and handle such mixed sentiment patterns, resulting in oversimplification or misclassification.
[70]	SentiCircles	OMD, HCR, STS-Gold	72.39	may not adequately capture sentiment expressions related to specific topics, events, or cultural aspects.
[71]	Unsupervised lexicon models	STS and OMD	72.6 for STS and 69.2 for OMD	Tweets frequently contain mixed sentiments or complex emotional expressions, which may not be accurately captured by unsupervised lexicon-based methods alone.

Additionally, Twitter-specific hybrid and ensemble methods for sentiment analysis are explored. When a large number of characteristics are used in the model, the study found that ML techniques, in particular SVM and MNB, deliver a high degree of accuracy. Although support vector machine (SVM) classifiers have become the de facto norm, lexicon-based approaches have proven to be effective with relatively low levels of human labour for tagging. Naive Bayes, Maximum Entropy, and Support Vector Machines all achieved around 80% accuracy when used in conjunction with n-gram and bigram models. Additionally, ensemble and hybrid-based algorithms performed better than supervised ML techniques (about 89%) when evaluating the sentiment of messages on Twitter.

Through a comprehensive review of sentiment analysis techniques in recent papers, we conducted an analysis focused on the classification algorithms, datasets used, and accuracy achieved. This analysis aimed to provide an overarching assessment of sentiment analysis on Twitter. By deconstructing numerous existing works, this review

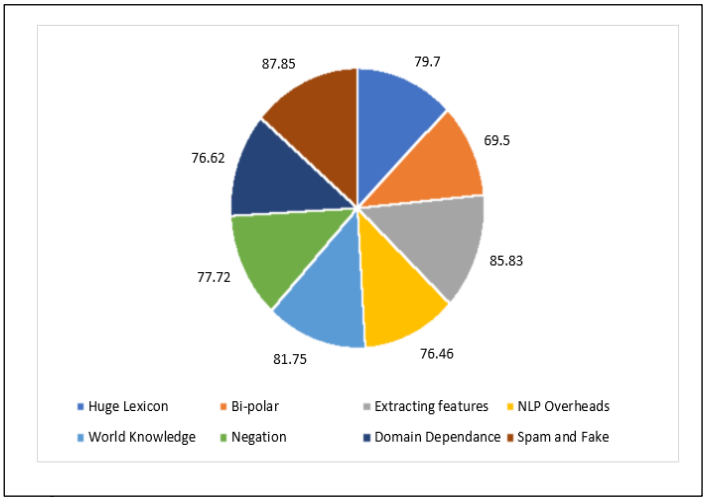


Fig. 17. Enhancing accuracy poses challenges in sentiment analysis

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Nistor, S. C., Moca, M., Moldovan, D., Oprean, D. B., & Nistor, R. L. (2021). Building a twitter sentiment analysis system with recurrent neural networks. *Sensors*, *21*(7), 2266.

[2] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," ACM Comput. Surv., vol. 49, no. 2, pp. 1-41, 2016.

[3] Radhi Desai, "Sentiment Analysis of Twitter Data : A Survey", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT) ,Volume , Issue 1-2456-3307, 2018.

[4] Twitter, Inc. Second Quarter 2016 Report. 2016. Retrieved from https://investor.twitterinc.com/results.cfm.

[5] S. Singh and P. Kumar, "Sentiment Analysis of Twitter Data: A Review," 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2023, pp. 1-7, doi: 10.1109/INOCON57975.2023.10100998.

[6] Keshavarz H, Abadeh M-S (2017) ALGA: adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs. Knowl-Based Syst 122:1–16.

[7] D. Goularas and S. Kamis, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019, pp. 12-17, doi: 10.1109/Deep-ML.2019.00011.

[8] Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of Indian farmers' protest using twitter data. International Journal of Information Management Data Insights, 1(2), 100019. https://doi.org/10.1016/j.jjimei.2021.100019.

[9] Behl, S., Rao, A., Aggarwal, S., Chadha, S., & Pannu, H. (2021). Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises. International Journal of Disaster Risk Reduction, 55, 102101. https://doi.org/10.1016/j.ijdrr.2021.102101.

[10] Tan, K. L., Lee, C. P., Lim, K. M., & Anbananthen, K. S. M. (2022). Sentiment Analysis With Ensemble Hybrid Deep Learning Model. IEEE Access, 10, 103694–103704. https://doi.org/10.1109/access.2022.3210182.

[11] Lu, Q., Zhu, Z., Zhang, D., Wu, W., & Guo, Q. (2020). Interactive Rule Attention Network for Aspect-Level Sentiment Analysis. IEEE Access, 8, 52505-52516,, https://doi.org/10.1109/ACCESS.2020.2981139.

[12] Mehta, K & Panda, S. (2019). A Comparative Analysis Of Sentiment analysis In Big Data. International Journal of Computer Science and Information Security, 17, 31-40.

[13] J He, J., Wumaier, A., Kadeer, Z., Sun, W., Xin, X., & Zheng, L. (2022). A Local and Global Context Focus Multilingual Learning Model for Aspect-Based Sentiment Analysis. IEEE Access, 10, 84135–84146. https://doi.org/10.1109/access.2022.3197218.

[14] E. Psomakelis, K. Tserpes, D. Anagnostopoulos, and T. Varvarigou, "Comparing methods for twitter sentiment analysis," KDIR 2014 - Proceedings of the Int. Conf. on Knowledge Discovery and Information Retrieval, pp. 225-232, 2014.

[15] Qurat Tul Ain_, Mubashir Ali_, Amna Riazy, Amna Noureenz, Muhammad Kamranz, Babar Hayat_ and A. Rehman, Sentiment Analysis Using Deep Learning Techniques: A Review , International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.

[16] A. Lopez-Chau, D. Valle-Cruz, and R. Sandoval-Almaz ´ an, "Sentiment ´ Analysis of Twitter Data Through Machine Learning Techniques," Software Engineering in the Era of Cloud Computing, pp. 185–209, 2020. Publisher: Springer, Cham.

[17] P. Kalaivani and D. Dinesh, "Machine Learning Approach to Analyze Classification Result for Twitter Sentiment," in 2020 International Conference on Smart Electronics and Communication (ICOSEC), (Trichy, India), pp. 107–112, IEEE, Sept. 2020.

[18] A. B. S, R. D. B, R. K. M, and N. M, "Real Time Twitter Sentiment Analysis using Natural Language Processing," International Journal of Engineering Research & Technology, vol. 9, July 2020. Publisher: IJERT-International Journal of Engineering Research & Technology.

[19] J. Ranganathan and A. Tzacheva, "Emotion Mining in Social Media Data," Procedia Computer Science, vol. 159, pp. 58–66, Jan. 2019.

[20] S. Xiong, H. Lv, W. Zhao, and D. Ji, "Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings," Neurocomputing, vol. 275, pp. 2459–2466, Jan. 2018.

[21] Y. M. Wazery, H. S. Mohammed, and E. H. Houssein, “Twitter Sentiment Analysis using Deep Neural Network,” in 2018 14th International Computer Engineering Conference (ICENCO), (Cairo, Egypt), pp. 177– 182, IEEE, Dec. 2018.

[22] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, “Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python,” International Journal of Computer Applications, vol. 165, pp. 29–34, May 2017. Publisher: Foundation of Computer Science (FCS), NY, USA.

[23] A. Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, “Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques.,” International Journal of Engineering and Technology, vol. 7, pp. 2038– 2044, Jan. 2016.

[24] P. FICAMOS and Y. LIU, “A Topic based Approach for Sentiment Analysis on Twitter Data,” International Journal of Advanced Computer Science and Applications, vol. 7, Dec. 2016.

[25] G. Gautam and D. Yadav, “Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis,” Sept. 2014.

[26] S. Ge, H. Isah, F. Zulkernine, and S. Khan, “A Scalable Framework for Multilevel Streaming Data Analytics using Deep Learning,” arXiv:1907.06690 [cs, eess], July 2019. arXiv: 1907.06690.

[27] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, “Deep Convolution Neural Networks for Twitter Sentiment Analysis,” IEEE Access, vol. 6, pp. 23253– 23260, 2018.

[28] S. Dhar, S. Pednekar, K. Borad, and A. Save, “Sentiment Analysis Using Neural Networks: A New Approach,” in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), (Coimbatore), pp. 1220–1224, IEEE, Apr. 2018.

[29] P. Vateekul and T. Koomsubha, “A study of sentiment analysis using deep learning techniques on Thai Twitter data,” in 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), (Khon Kaen, Thailand), pp. 1–6, IEEE, July 2016.

[30] Singh, Prabhsimran, Ravinder Singh Sawhney, and Karanjeet Singh Kahlon. "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government." ICT Express (2017).

[31] J. F. Raisa, M. Ulfat, A. Al Mueed and S. M. S. Reza, "A Review on Twitter Sentiment Analysis Approaches," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2021, pp. 375-379, doi: 10.1109/ICICT4SD50815.2021.9396915.

[32] Gupta, Bhumika, et al. "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python." International Journal of Computer Applications 165.9 (2017).

[33] Adwan OY, Al-Tawil M, Huneiti AM, Shahin RA, Abu Zayed AA, Al-Dibsi RH (2020) Twitter sentiment analysis approaches: a survey. Int J Emerg Technol Learn. https://doi.org/10.3991/ijet.v15i15.14467.

[34] A. Barhan and A. Shakhomirov, "Methods for Sentiment Analysis of twitter messages," in 12th Conference of FRUCT Association, 2012.

[35] A. Krouska, C. Troussas and M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis," 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), Chalkidiki, Greece, 2016, pp. 1-5, doi: 10.1109/IISA.2016.7785373.

[36] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.

[37] M. Day and C. Lee, Deep Learning for Financial Sentiment Analysis on Finance News Providers, no. 1, pp. 11271134, 2016.

[38] M. Cliche, “Bb_twtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms,” arXiv Prepr. arXiv1704.06125, 2017.

[39] M. Thomas and L. C.A, “Sentimental analysis using recurrent neural network,” Int. J. Eng. Technol., vol. 7, no. 2.27, p. 88, 2018.

[40] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, and E. Cambria, “Learning Word Representations for Sentiment Analysis,” Cognit. Comput., vol. 9, no. 6, pp. 843–851, 2017.

[41] K. Baktha and B. K. Tripathy, “Investigation of recurrent neural networks in the field of sentiment analysis,” Proc. 2017 IEEE Int. Conf. Commun. Signal Process. ICCSP 2017, vol. 2018–January, pp. 2047–2050, 2018.

[42] P. Liu, X. Qiu, and H. Xuanjing, “Recurrent neural network for text classification with multi-task learning,” IJCAI Int. Jt. Conf. Artif. Intell., vol. 2016–January, pp. 2873–2879, 2016.

[43] A. Hassan and A. Mahmood, “Deep Learning approach for sentiment analysis of short texts,” 2017 3rd Int. Conf. Control. Autom. Robot. ICCAR 2017, pp. 705–710, 2017.

[44] Ruangkanokmas, P.; Achalakul, T.; Akkarajitsakul, K. Deep Belief Networks with Feature Selection for Sentiment Classification. In Proceedings of the 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, Thailand, 25–27 January 2016; pp. 9–14.

[45] Socher, R.; Lin, C.C.; Manning, C.; Ng, A.Y. Parsing natural scenes and natural language with recursive neural networks. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 129–136.

[46] Long, H.; Liao, B.; Xu, X.; Yang, J. A hybrid deep learning model for predicting protein hydroxylation sites. Int. J. Mol. Sci. 2018, 19, 2817. [CrossRef].

[47] Vateekul, P.; Koomsubha, T. A study of sentiment analysis using deep learning techniques on Thai Twitter data. In Proceedings of the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 13–15 July 2016; pp. 1–6.

[48] Ghosh, R.; Ravi, K.; Ravi, V. A novel deep learning architecture for sentiment classification. In Proceedings of the 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 3–5 March 2016; pp. 511–516.

[49] Available online: http://help.sentiment140.com/site-functionality (accessed on 12 March 2020).

[50] Available online:https://www.kaggle.com/crowdflower/twitter-airline-sentiment (accessed on 12 March 2020 Fang, X., Klawohn, J., De Sabatino, A., Kundnani, H., Ryan, J., Yu, W., & Hajcak, G. (2022). Accurate classification of depression through optimized machine learning models on high-dimensional noisy data. Biomedical Signal Processing and Control, 71(PartB),103237. https://doi.org/10.1016/j.bspc.2021.103237.

[51] L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learninig Methods," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154183.

[52] Saleena, Nabizath. "An ensemble classification system for twitter sentiment analysis." *Procedia computer science* 132 (2018): 937-946.

[53] K. N. Alam et al., Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data, Comput. Math. Methods Med. 2021 (2021) e4321131. doi:10.1155/2021/4321131.

[54] I. Krak, O. Barmak, P. Radiuk, Information technology for early diagnosis of pneumonia on individual radiographs, in 3rd International Conference on Informatics & Data-Driven Medicine (IDDM-2020) 2753 (2020) 11–21. [Online]. Available: http://ceur-ws.org/Vol-2753/paper3.pdf.

[55] H. Jang, E. Rempel, D. Roth, G. Carenini, N. Z. Janjua, Tracking COVID-19 discourse on Twitter in North America: Infodemiology study using topic modelling and aspect-based sentiment analysis, J Med. Internet Res. 23(2) (2021) e25431. doi:10.2196/25431.

[56] G. Yenduri, B. R. Rajakumar, K. Praghash, D. Binu, Heuristic-assisted BERT for Twitter sentiment analysis, Int. J. Comput. Intell. Appl. 20(03) (2021) e2150015. doi:10.1142/S1469026821500152.

[57] A. Addawood et al., Tracking and understanding public reaction during COVID-19: Saudi Arabia as a use case, Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020 24(1) (2020) 1–9. doi:10.18653/v1/2020.nlpcovid19-2.24.

[58] N. S. Sattar, S. Arifuzzaman, COVID-19 vaccination awareness and aftermath: Public sentiment analysis on twitter data and vaccinated population prediction in the USA, Appl. Sci. 11(13) (2021) e6128. doi:10.3390/app11136128.

[59] Radiuk, Pavlo, Olga Pavlova, and Nadiia Hrypynska. "An ensemble machine learning approach for Twitter sentiment analysis." (2022).

[60] H. A. Shehu and S. Tokat, “A Hybrid Approach for the Sentiment Analysis of Turkish Twitter Data,” in Artificial Intelligence and Applied Mathematics in Engineering Problems (D. J. Hemanth and U. Kose, eds.), Lecture Notes on Data Engineering and Communications Technologies, (Cham), pp. 182–190, Springer International Publishing, 2020.

[61] A. Kumar and G. Garg, “Sentiment analysis of multimodal twitter data,” Multimedia Tools and

Applications, vol. 78, pp. 24103–24119, Sept. 2019.

[62] N. Mittal, B. Agarwal, S. Agarwal, S. Agarwal, and P. Gupta, “A Hybrid Approach for Twitter Sentiment Analysis,” 2015.

[63] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, “Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis,” p. 8.

[64] M. M. Fouad, T. F. Gharib, and A. S. Mashat, "Efficient Twitter Sentiment Analysis System with Feature Selection and lassifier Ensemble," in International Conference on Advanced Machine Learning Technologies and Applications, 2018: Springer, pp. 516-527.

[65] Zahoor S, Rohilla R (2020) Twitter sentiment analysis using lexical or rule based approach: a case study. In: ICRITO 2020—IEEE 8th international conference on reliability, Infocom technologies and optimization (trends and future directions). https://doi.org/10.1109/ICRITO48877.2020.9197910.

[66] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, "Lexiconenhanced sentiment analysis framework using rule-based classification scheme," PloS one, vol. 12, no. 2, p. e0171649, 2017.

[67] F. M. Kundi, A. Khan, S. Ahmad, and M. Z. Asghar, "Lexicon-based sentiment analysis in the social web," Journal of Basic and Applied Scientific Research, vol. 4, no. 6, pp. 238-48, 2014.

[68] X. Hu, J. Tang, H. Gao, and H. Liu, “Unsupervised sentiment analysis with emotional signals,” in Proceedings of the 22nd international conference on World Wide Web, WWW ’13, (New York, NY, USA), pp. 607–618, Association for Computing Machinery, May 2013.

[69] H. Saif, Y. He, M. Fernandez, and H. Alani, “Contextual semantics for sentiment analysis of Twitter,” Information Processing and Management: an International Journal, vol. 52, pp. 5–19, Jan. 2016.

[70] X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised sentiment analysis with emotional signals," in Proceedings of the 22nd international conference on World Wide Web, 2013: ACM, pp. 607-618.

[71] S. Singh and P. Kumar, "Sentiment Analysis of Twitter Data: A Review," 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2023, pp. 1-7, doi: 10.1109/INOCON57975.2023.10100998.

[72] Ali, Kashif, Hai Dong, Athman Bouguettaya, Abdelkarim Erradi, and Rachid Hadjidj. (2017) “Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework”, in IEEE International Conference on Web Services (ICWS), Honolulu, HI, USA: IEEE.

[73] Ragini, J. Rexiline, P. M. Rubesh Anand, and Vidhyacharan Bhaskar. (2018) “Big Data Analytics for Disaster Response and Recovery Through Sentiment Analysis.” International Journal of Information Management 42: 13-24.

[74] Hassan, Anees Ul, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, and Sungyoung Lee. (2017) “Sentiment Analysis of Social Networking Sites (SNS) Data Using Machine Learning Approach for the Measurement of Depression”, in International Conference on Information and Communication Technology Convergence (ICTC), Jeju, South Korea: IEEE.

[75] Joyce, Brandon, and Jing Deng. (2017) “Sentiment Analysis of Tweets for the 2016 US Presidential Election”, in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE.

[76] Martin-Domingo, Luis, Juan Carlos Martin, and Glen Mandsberg. (2019) “Social Media as a Resource for Sentiment Analysis of Airport Service Quality (ASQ).” Journal of Air Transport Management.

[77] Isah, Haruna, Paul Trundle, and Daniel Neagu. (2014) “Social Media Analysis for Product Safety Using Text Mining and Sentiment Analysis”, in 14th UK Workshop on Computational Intelligence (UKCI): IEEE.

[78] Akter, Sanjida, and Muhammad Tareq Aziz. (2016) “Sentiment Analysis on Facebook Group Using Lexicon Based Approach”, in the 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT).

[79] Twitter Users: How Many People Use Twitter in 2023? (NEW Stats) - EarthWeb.

[80] Dang NC, Moreno-García MN, De la Prieta F. Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*. 2020; 9(3):483. https://doi.org/10.3390/electronics9030483.

[81] Lopez, Marc, and Kalita, Jugal. "Deep Learning Applied to NLP." *ArXiv*, 2017, /abs/1703.03091. Accessed 23 Jul. 2023.