Exploring Sentiment and Emotion Analysis: A Systematic Review and Future Directions

Article *in* International Journal of Electrical and Electronics Engineering · December 2024

DOI: 10.14445/23488379/IJEEE-V11I12P107

Nisha Hooda
Kurukshetra University

SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I12P107

Volume 11 Issue 12, 76-93, December 2024

Review Article

Exploring Sentiment and Emotion Analysis: A Systematic Review and Future Directions

Nisha[1], Rakesh Kumar[2]

[1,2]*Department of Computer Science & Applications, Kurukshetra University, Haryana, India.*

[1]*Corresponding Author : nisha.jrf.dcsa@kuk.ac.in*

Received: 05 October 2024 Revised: 06 November 2024 Accepted: 04 December 2024 Published: 31 December 2024

Abstract - *Human communication often carries emotional undertones expressed verbally, through written texts, and non-vocally via gestures and facial expressions. With the widespread adoption of texting and the rapid growth of social media, digital communication has become more prevalent. The Internet is crucial in the modern era of technology, providing a vast repository of information that necessitates thorough examination to extract relevant insights. Sentiment analysis, which involves computationally analyzing sentiments, viewpoints, and the subjective aspects of text, is crucial for uncovering hidden information and accurately classifying emotions. However, obtaining suitable datasets and applying precise classifiers pose significant challenges. The function of artificial intelligence technology in automated text sentiment analysis is the primary subject of this article, which also examines other uses, difficulties, and approaches to sentiment analysis. Drawing attention to obstacles, constraints, and potential avenues for further study adds to what is already known and helps academics and professionals choose the best approaches.*

Keywords - *Emotion detection, Machine Learning, Natural Language Processing, Sentiment classification, Text analysis.*

1. Introduction

The term "sentiment" has its roots in the French words "santement, sentement," which translates to "feeling, affection, and opinion”. The significance of sentiment became prominent following the advancements in cognitive psychology during the mid-twentieth century, as cognitive theories illustrated the essential role of emotion and thought in our daily lives.

Proper comprehension of these emotions by a machine to create a supportive environment for Human-Computer Interaction (HCI) remains a highly debated academic topic. Twitter and Facebook, popular social media platforms, are extensively utilized by individuals worldwide to articulate their thoughts and opinions on various subjects. Identifying human apprehensions in context, such as customer feedback about new products, film critiques, and responses to government legislation, is crucial.

Examining the sentiments expressed in textual content on digital platforms proves immensely advantageous for this objective. Individuals tend to use informal methods of expression more frequently on social media platforms than formal ones, resulting in several challenges regarding its categorization. Opinion mining or Sentiment Analysis (SA) of social media posts involves determining the viewpoint, attitude, or emotions expressed in these posts [1]. These sentiments can be discussed at several levels and are primarily classified as positive and negative. It has advanced even further to discern the more subtle degrees of emotions in later phases.

The purpose of emotion detection and classification is to identify more intricate emotions, such as happiness, sadness, and rage, if they are present in the sentence. Emotion recognition refers to the process of identifying and discerning more nuanced levels of emotions. Nowadays, nearly every area of company or industry is experiencing a digital revolution, resulting in a vast quantity of organized and unstructured data. Converting disorganized data into valuable knowledge for decision-making poses a significant challenge for businesses [2].

Twitter serves as a critical source of health-related information, particularly amidst the COVID-19 pandemic, with a substantial number of headlines and views being shared. Using topic modeling, relevant content was extracted by analyzing a large English and Portuguese tweets dataset. This analysis aimed to comprehend human behaviors in affected areas and the correlation between news announcements and mood over four months [3]. Sentiment assessment has transformed business operations by providing marketers with insights into customer perspectives and facilitating necessary adjustments. A study conducted on the

SSE 50 Index in China demonstrated that combining sentiment pointers with share market data led to enhanced performance [4]. Utilizing the supervised ML algorithm Support Vector Machine (SVM) and a rolling window approach resulted in a notable 18.6% increase in precision. The suggested method improves data processing efficiency by analyzing sentiment and emotions in different settings using lexical analysis, Machine Learning (ML), and Deep Learning (DL).

Although they possess distinct characteristics, each has its own benefits and drawbacks. Researchers continue to face significant challenges in this field, including handling context, ridicule, utterances with various emotions, the emergence of digital jargon, and vagueness in grammar and vocabulary. Furthermore, because there are a large number of resources accessible in widely spoken dialects such as English, most investigations for the identification of emotions focus on analyzing data from a single language. However, when utilizing social media, most individuals proficient in many languages often engage in code-mixing [5].

Furthermore, due to the absence of established protocols for conveying emotions through different channels, certain individuals are able to effectively transmit their emotions, while others suppress them and maintain a rational presentation of their message. This study seeks to address this requirement by examining the present condition, difficulties, and prospective course in emotion detection and SA, specifically emphasising ML-based methodologies for emotion detection and SA. This review mainly aims to answer the following research question:

- Q1-Which ML methods are employed for subjectivity detection?
- Q2-Which ML algorithms yield the highest performance in sentiment detection?
- Q3-When trying to resolve the issue of emotion detection and evocation, what data sources are used?
- Q4-What are the possible roadblocks, knowledge gaps, and research directions to take next?

A systematic literature review of 101 articles was done to answer these questions. The distribution of publications over the years is considered for this research, as shown in Figure 1. The report raises many difficulties based on the findings and review. The challenges in this field include 1) a dearth of text-based datasets for languages apart from English, 2) a lack of standardized datasets and evaluation procedures that impede the ability to reproduce experiments, 3) a limited emphasis on non-textual data sources due to the unavailability of suitable datasets, and 4) the disregard of factors such as processing and learning time in model evaluation, which are essential for real-time interactive systems. The report highlights the potential for future research to tackle these problems, thereby promoting the advancement of interactive systems design and facilitating the creation of emotionally engaging systems.

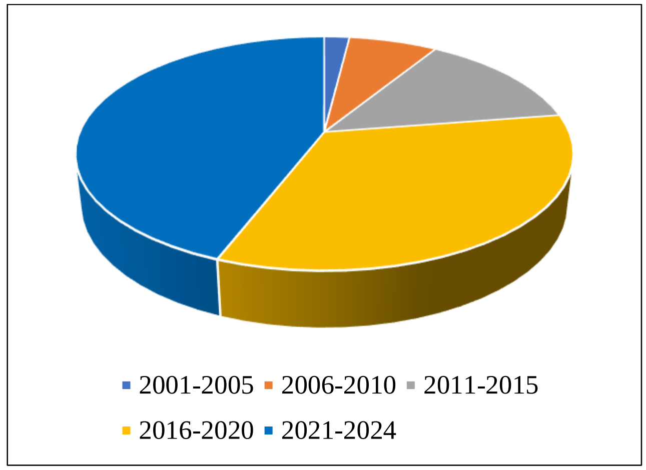

Fig. 1 Distribution of publications

The outline of the rest of the paper is organized as follows. Comprehensive information about the background of SA and its many degrees is presented in Section 2. Section 3 covers the many phases of the sentiment and emotion analysis process, which include datasets, text preprocessing, feature extraction, and selection approaches. Section 4 examines the diverse obstacles encountered in the realm of SA. Section 5 provides an overview of the methodologies used to conduct SA. Section 6 contains a comprehensive analysis of model evaluation measures. The survey concludes in Section 7.

2. Related Works

This part gives a synopsis of the primary principles and initial information relevant to this study. More precisely, it presents the notions of SA and various levels of doing it.

2.1. Sentiment and Emotion Analysis

Today, a significant proportion of individuals utilize discussion boards, weblogs and various digital media podiums such as YouTube, Twitter, and Instagram to communicate individual ideas and opinions to a global audience. This leads to generating a substantial volume of information called big data. SA was designed to examine this vast amount of data highly efficiently and effectively [6].

An industry or corporation must possess the capacity to know the emotions of the user. SA is the practice of collecting pertinent information from written material using Natural Language Processing (NLP) methods and determining the authors' sentiment, which can be favorable, adverse, or neutral. The primary aim of SA is to detect subjectivity and categorization in biased writing, typically as either positive or negative. However, it can also encompass descriptors such as pleasant or dreadful, agree or disapprove, and neutral [7]. Subsequent paragraphs delve into various studies in this area.

A study [8] highlighted the significance of how ML practices could automate SA for the first time. Unlike traditional text categorization, which relied heavily on keywords to identify subjects, this research study showed that expressing thoughts more nuancedly posed greater challenges. The study conducted a comparison between movie reviews and author evaluations and discovered that sentiment categorization can be accomplished by utilizing positive and negative groupings. The SVM, Maximum Entropy, and Naive Bayes (NB) classifiers demonstrated a 90% accuracy rate in conventional topic-based classification.

Turney's [9] unsupervised ML methodology classifies text reviews as positive or negative, using the mean semantic orientation of adjectives or adverbs. The study collected 410 reviews across four domains: vehicles, banking, movies, and vacation places. The method achieved an accuracy of 74%, the highest in the vehicle dataset. A comparison was made amongst famous supervised ML algorithms, namely NB, SVM, and the character-based N-gram approach [10], to categorize travel and tourism blog sentiment for seven US and European vacation sites. Results showed that SVM and N-gram algorithms outperformed NB, with negative reviews.

A diverse multiclass Urdu dataset was created by collecting user reviews from many businesses, such as food and beverage, entertainment, technology, politics, and sports. The proposed dataset consists of 9312 reviews, which have been meticulously categorized into three classifications, positive, negative, and neutral, by human professionals. The primary goal of this project is to provide a manually annotated dataset for Urdu SA.

Additionally, the research aims to produce baseline results by employing rule-based, ML, and DL methods [11]. NLP is an academic field that investigates the relationship between machines and human languages. One prominent application within this topic is opinion mining. The objective of SA of Hindi (SAH) in India is to ascertain the polarity of Hindi sentences by utilizing Hindi stop words and the Natural Language Toolkit (NLTK). The study employs DT and NB Classifiers to collect data and extract text [12].

2.1.1. Stages of SA

When initiating SA, the primary step involves choosing the right degree of abstraction. Throughout the SA process, three distinct levels of abstraction may be encountered, namely document level, sentence level, and feature level, as illustrated in Figure 2. There is a very low degree of detail at the document level, and the amount of detail gradually increases in a simulated fashion up to the aspect level. The primary objective of document-level SA is to ascertain the document's overall polarity. This type of SA is not commonly used in modern times. This approach enables the classification of individual paragraphs within a novel as positive, negative, or neutral. It's imperative to distil the overarching sentiment from lengthy texts amidst the noise and replicate localized trends.

Expressing opinions through text presents a challenge due to the diverse and often conflicting viewpoints, which can be conveyed implicitly, highlighting the intricate relationships among words, sentences, and overall contextual semantics to reflect document structure. This level necessitates a comprehensive understanding of the nuanced structural patterns of sentiment and their dependencies on words [13].

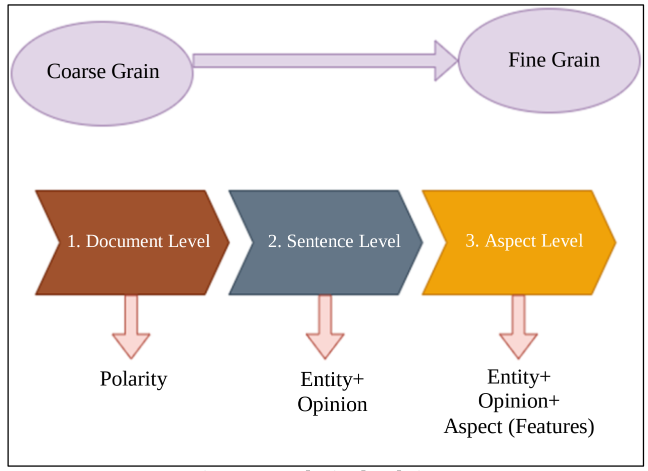

Fig. 2 Granularity levels in SA

The sentiment of each sentence is considered in the categorization at the phrase or sentence level, which is very valuable for identifying subjectivity. Generally, statements in written writings either express subjective viewpoints or do not. The process of subjective classification involves determining whether a statement in a document consists of factual information, emotional expressions, or personal opinions. The main goal of subjectivity classification is to eliminate sentences that do not express an opinion. For instance, when someone says, "I received a stunning card," it indicates that the card has a good sentiment. Hence, it's regarded as a subjective assertion that can be subdivided into multiple polarities. Nonetheless, the assertion "This bag is red in color" lacks any emotional connotation, thus falling under the category of an objective statement. Handling conditional or vague expressions in sentence-level SA holds significant importance.

The aspect level is the most detailed level of SA, where each aspect mentioned in the sentence is carefully evaluated and assigned a polarity before an overall sentiment is determined for the whole phrase. Identifying implicit traits in opinion mining, particularly at this level, is more complicated than the above two levels [14]. For instance, the car has excellent fuel efficiency but is expensive. Here, mileage and cost are two perspectives. The phrase expressly mentions mileage as a specific component, but cost is implied as a feature in the sentence. While phrases and documents can be utilized for sentiment research, these alternatives are regrettably inadequate for detailed SA. Concentrating on aspect-driven SA has great potential for accurately predicting the sentiments expressed by users. An innovative approach is introduced to extract implicit aspects from opinionated writings. This algorithm rates each pair of aspects and feature indicators based on their frequency of co-occurrence [15]. Figure 3 shows the distribution over the granularity of text data.

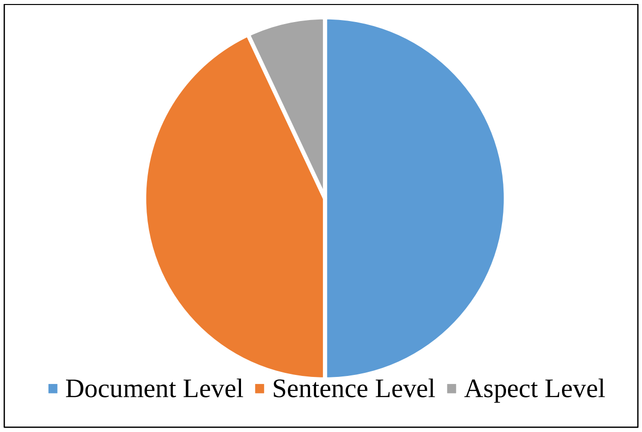

Fig. 3 Distribution over granularity of text data

In addition to the three primary levels of SA, known as basic levels, there is a subset of sentiment levels referred to as sub-level SA. Multidomain and multimodal sentiment categorization involve transferring information across domains, thereby advancing affective computing. This area gains significance as individuals increasingly express their thoughts through videos and images on social media platforms [16]. Multimodal approaches necessitate the integration of features from various modalities, employing early or late fusion techniques. Such generalization is pivotal for opinion mining and SA within the multimedia community [17].

3. SA Process

To uncover opinions from the provided data, several steps outlined in Figure 4 must be adhered to. Subsequent paragraphs elucidate the entire process of extracting sentiments from a given text.

3.1. Data Collection

Data collection is the preliminary stage of situational awareness. Data can be collected via the Internet through various means such as web mining, social networking platforms, media outlets, e-commerce sites, online forums, and weblogs. Textual data can be integrated with various other data streams, including visual, audio, and geolocation, depending on the objective of the analysis [18]. This data can be acquired from several significant sources. An internet forum, also referred to as a message board, is a platform where users may engage in text-based discussions, share their experiences, ask questions, and exchange ideas on topics that interest them [19]. Forums are an attractive resource for SA due to the dynamic nature of user-generated content found inside them [20].

Digital Platform: In the contemporary era, the ability to connect to the internet is an essential commodity. Individuals are increasingly relying on the applications provided by social networking platforms to articulate their viewpoints on contemporary issues. Over the past few decades, there has been a rapid increase in the use of SA to assess customer sentiments through social media platforms [21]. A weblog is a type of website that publishes information in reverse chronological order. The platform functions as a medium for writers to articulate their viewpoints on particular subjects, making it a valuable instrument for doing SA on various entities [22]. E-commerce websites are platforms where users may provide reviews and express their thoughts about a certain firm or organization, including product reviews or expert evaluations.

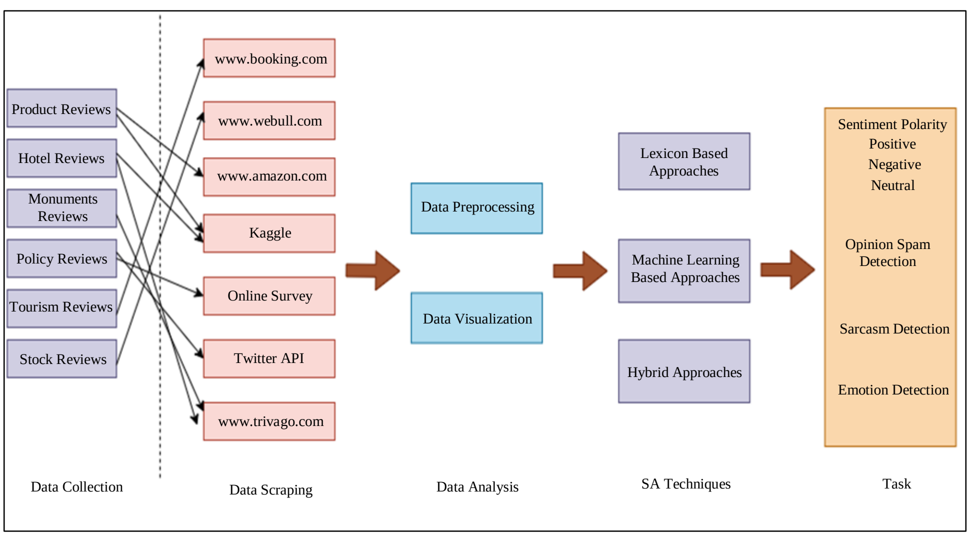

Fig. 4 General process of SA

The data collected from the aforementioned sources is unstructured and has to be reformatted for further analysis. Within the field of literature, numerous established and categorized datasets exist that are commonly used for the SA process. These datasets may also necessitate the implementation of feature engineering techniques. The data required for the SA process should be specific to the domain, extensively documented, and substantial in quantity. SA datasets are characterized by their preprocessing requirements, format, and quantity of occurrences. Sentiment 140, US airlines, IMDB, and Yelp are some often used datasets for the process of textual SA. The distribution of publications over domain/area is shown in Figure 5.

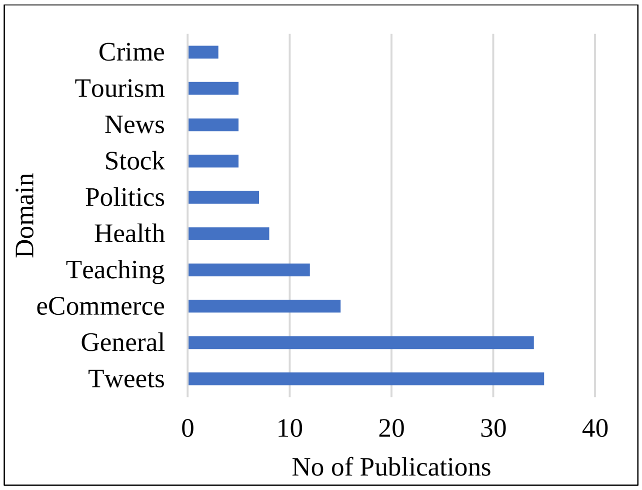

Fig. 5 Distribution over domain/area

Bouazizi et al. [23] introduced a novel approach to identifying sarcasm by analyzing the emotion of Twitter data. They highlight the effectiveness of microblogging social networking sites in detecting sarcastic statements. The sarcasm detection for Twitter uses a pattern-based technique. A feature set consisting of four pertinent elements is employed to distinguish between various forms of sarcasm. Tweets are then categorized into two classes: non-sarcastic and sarcastic. The model attained an accuracy rate of 83.1% and a precision rate of 91.1%. The implementation utilizes the SVM classifier and the WEKA tool. Figure 6 shows the dataset distribution commonly employed for textual SA.

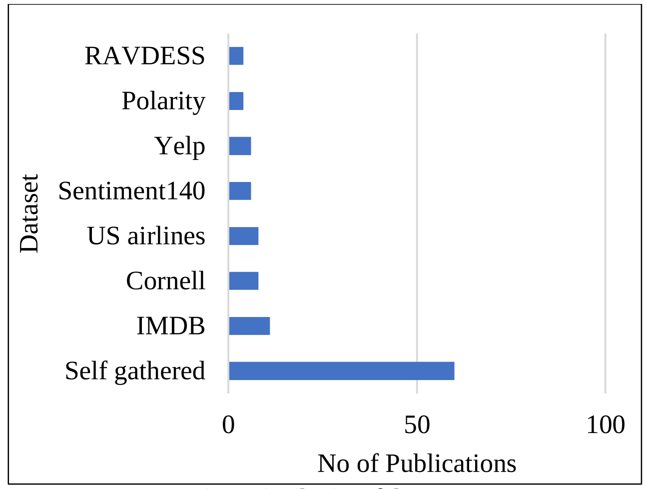

Fig. 6 Distribution of datasets

3.2. Data Preprocessing

Previous investigations have utilized a range of preprocessing approaches [24]. The datasets that have been gathered are in their original, unprocessed format. As a result, different preprocessing techniques yield different levels of accuracy when it comes to categorization. In order to ascertain the most effective approach, multiple experiments were carried out, employing different methodologies or a combination of procedures. Several toolkits can be used for NLP tasks and text preparation. For example, NLTK [25] in Python, Texblob in Python [26], Open NLP [27], Core NLP [28], and MADAMIRA [29] in Java. Pre-processing is essential for accurate prediction [30] as it decreases processing overhead and enables more effective training of ML models. Not every situation can be solved with a single solution. The methods employed for preparing textual data in SA are presented in Table 1.

Table 1. Various techniques for preprocessing

No	Methods	Explanation
1	Normalization [31]	In order to achieve normalization, multiple activities are performed simultaneously, including altering the text to lowercase, removing URLs, and eliminating punctuation, hashtags, and whitespace. This enhances the uniformity of the preparation process for each text.
2	Tokenization [32]	The process of dividing a sequence of text into smaller pieces, formally known as tokens, so is the process called as tokenization. Some tools available for tokenizing a text document are NLTK's Word tokenize and TextBlob.
3	Stemming [33]	The most frequent approach for eliminating prefixes, suffixes, and definite articles from words in order to return them to their original root form. Root stemming, mild stemming, and hybrid stemming are some of the stemming methods that are offered.
4	Lemmatization [34]	It involves the merging of two or more words into a single word. Through morphological examination, the termination of the word is eliminated.

5	Removal of Stop Words [35]	These terms are the most often used in any language and are irrelevant to SA. For instance, in English, common stop words include "at," "is," and "and," among others.
6	Amending Abbreviations and Informal Language [36]	The process involves transforming abbreviations into their full-word equivalents to enhance the comprehension of the term by machines. Before: TBT to our trip to NYC! After: Throwback Thursday to our trip to New York City!

3.3. Feature Selection and Feature Extraction

In textual data processing, a feature is commonly depicted as a vector known as a feature vector. In SA, feature extraction and feature selection are utilized to manage these features.

Feature extraction refers to the conversion of textual input into a numerical form and the subsequent selection of the most significant features [37]. The process of representing words numerically is commonly referred to as word embedding. By generating additional features that extract the most important information from the initial collection of features, feature extraction algorithms improve the performance of machine learning models [38]. Several textual feature representation strategies have been previously discussed, each possessing distinct characteristics [39]. Count-based and optimization-based statistical methods, for instance, rely on prediction.

The count approach relies on determining the word count of the document. If the phrase appears in the text, it is assigned a value of one; otherwise, it is allocated a zero value. One-hot encoding and Bag-of-Words (BoW) are techniques used in NLP and ML. Term Frequency (TF), Inverse Document Frequency (IDF), and Co-occurrence matrix are widely used techniques for word embedding based on counting [40]. Word2vec, Continuous Bag of Words (CBOW), Global Vectors (GloVe) [41], and skip-gram are widely used algorithms for the prediction method. In general, it is highly compatible with DL techniques. An essential aspect of the feature extraction process is selecting a highly efficient methodology for representation. There are multiple methods and different degrees of depicting characteristics [42].

Feature selection involves identifying data qualities that are significant, trivial, or redundant using a specific approach. The objective of feature selection approaches in ML is to identify the ideal set of characteristics that may be utilized to develop an optimized model for a given job. Various feature selection approaches are employed to eliminate superfluous and inconsequential features.

The objective of choosing the most pertinent features from the original collection of features is to enhance the forecast's accuracy or reduce the structure's size by eliminating unnecessary qualities. One way to accomplish this is by selecting a subset of features from the original set while maintaining the accuracy of the classifier's predictions based on the selected features [43]. Several recent studies have examined methods for reducing feature vectors by excluding irrelevant characteristics, which can expedite calculations and enhance classification accuracy [44, 45].

The GAWA feature selection method, which uses Wrapper Approaches and Genetic algorithms, has been introduced for hybrid sentiment classification. This innovative approach minimizes redundancy, enhances accuracy, and can reduce feature sets by up to 61.95%. With an accuracy percentage of 92, NB outperformed all other approaches. In this study, wrapper techniques were employed to extract 8243 prime feature sets from Twitter data. These feature sets were then further condensed to 3137 ideal feature sets using the Genetic algorithm. The GAWA approach exhibits an 11% increase in accuracy compared to PCA and an 8% increase compared to PSO [46].

Feature extraction is the act of obtaining pertinent information from unorganized text by utilizing different feature selection approaches. Feature selection procedures can be categorized into two types: supervised and unsupervised. Supervised methods consider the target class variables, whereas unsupervised methods concentrate on discovering associations between input variables to minimize redundant variables [47, 48]. These methods are efficient and thorough in recognizing phrases that indicate the development of a limited range of characteristics. It is possible to pick and reduce features using a number of statistical feature selection methods, such as:

Filter approach: The filter method is a supervised technique that evaluates the intrinsic properties of features to assess their significance. Employing statistical methods such as ANOVA and chi-square, among others, sifts through features to identify highly relevant ones. This method operates swiftly and independently of a classifier; however, it overlooks the influence of the classification algorithm. In this study, feature selection is conducted using information gain and Chi-square techniques [49, 50].

The wrapper-based feature selection method adopts a greedy approach, striving to identify the optimal feature subset from input variables through an ML classifier [51]. It generates multiple models utilizing various subsets of input features and selects the feature set associated with the highest-performing model. Although this approach incurs higher costs than filtering methods, it excels in class prediction efficiency [52]. Common strategies include forward selection, backward elimination, and recursive feature elimination, all aimed at enhancing model efficiency and accuracy by iteratively adding and removing weaker features from the input feature set. Embedded feature selection methods integrate filter and wrapper-based approaches, enabling feature interaction while controlling computational expenses [53]. Utilizing techniques such as Decision Trees, LASSO, and RIDGE regression, these methods incorporate inherent learning algorithms and penalization functions to mitigate overfitting. They provide advantages such as feature interaction, speed, efficiency, and decreased overfitting.

4. Challenges in SA

SA poses several challenges due to computational costs, informal language with abbreviations, slang and grammar errors, alongside linguistic diversity, limiting ML models' effectiveness. Different sentiment structures further complicate the analysis, ranging from structured reviews like academic papers to semi-structured reviews. The latter presents a mix of formal and informal elements. Unstructured sentiments, conversely, involve spontaneous writing with implicit and explicit features. Implicit characteristics in text pose initial hurdles in analysis.

Mockery Detection: The SA process should necessitate a comprehension of how to manage circumstances that have several interpretations and the use of irony. For instance, sarcasm conveys a pessimistic attitude towards an item, yet conventional SA algorithms frequently fail to detect it. Various methodologies have been proposed for the identification of sarcasm in all languages. Sarcasm identification can be accomplished by either tag-based administration or using minor datasets annotated by humans [54].

Computational expense: In order to enhance precision, it is necessary to raise both the complexity of the model and the volume of the training data. Training the model with a huge corpus will significantly increase the computational cost, necessitating the use of a powerful GPU to achieve improved outcomes.

Insufficient resources: In order to accurately determine attitudes, certain statistical techniques require a substantial dataset that has been annotated. While data collection may not be challenging, manually annotating a large dataset requires significant effort and is prone to lower levels of accuracy. In addition, the dataset now accessible consists mainly of English languages, with additional languages yet to be included [55]. Detecting thoughts in languages other than English poses a significant difficulty but also presents exciting prospects for new research endeavors.

Informal Language: Individuals often compose reviews or texts in a highly casual manner, frequently employing acronyms, emoticons, and shortcuts that can impede comprehension of the material. Due to their lack of universality, acronyms cannot be accurately managed as they may vary by location and undergo rapid changes. Grammatical mistakes are frequently found in informal writing and can be somewhat repaired. However, these corrections have limitations that can hinder the accuracy of the analyzer. Additional issues involve managing code mix reviews, navigating multi-opinionated evaluations, and adapting writing styles for reviews due to linguistic variations across different regions and locales [56].

For example, the words "colour" and "color" have identical meanings but are spelt differently in various regions of the world. Analyzing accurate thoughts poses an additional issue when slang is used on several social media platforms. Existing lexicons and trained models are encountering a substantial obstacle due to the growing collection of internet slang terms [57].

5. Methodologies for SA

There are several considerations while deciding on sentiment analysis methods, such as the specific task at hand. Traditional ML methods are well suited for tasks like subjectivity analysis, sentiment sorting, and implicit language identification. Lexicon approaches are suitable for tasks involving sentiment, object, and aspect extraction. Hybrid techniques, however, are less effective in opinion detection tasks. DL is commonly used for sentiment classification tasks involving complex and non-textual data. Historically, emotion identification relied on lexicon-based analysis, ML, and hybrid approaches. Yet, as data complexity grows, novel techniques like DL and transfer learning become crucial. The process of SA is illustrated in Figure 7, with ongoing research aimed at enhancing precision and reducing computational costs. The paper primarily focuses on ML-based SA methodology.

5.1. Lexicon-Based Method

The traditional method relies on prebuilt lexicons containing lexical items with assigned sentiment values. Handcrafted features are also utilized here [58]. In the study [59], sentiment normalization and evidence-built combination roles were used to quantify the intensity of sentiment. This approach is more accurate than using basic aggregate or average functions, as demonstrated by previous research. This approach does not necessitate the acquisition of new information, as the tokens are predetermined in a pre-existing vocabulary.

This characteristic greatly enhances its suitability for analyzing sentiment at the phrase and aspect levels. The manual method alone is highly time-consuming and is not employed in all instances. This strategy incorporates the following automated techniques to ensure accuracy as a final verification step. The approach incorporates both a dictionary-based and a corpus-based method. The dictionary-based method manually gathers seed words and their polarities, refining them with antonyms and synonyms. Iterative adjustments and manual validations ensure accuracy, yet it struggles with identifying distinct subjective terms. Corpus-based methods prove more efficient with clear domains, although compiling a comprehensive English word collection poses challenges. Combining both approaches is less efficient due to this complexity. Table 2 highlights the pros and cons of the lexicon-based method.

Table 2. Overview of the advantages and disadvantages of the lexicon-based classification methodology

Methodology	Pros	Cons
Dictionary Method	Positive outcomes were observed within the narrow domain. There is no necessity for training data.	Identifying subjective terms that apply to various situations and fields is a difficult task. The approach's effectiveness is diminished as it fails to consider the contextual nuances of the emotional expression.
Corpus Method	Opinion words with context-dependent orientations can be more easily identified. Optimal results are achieved when domains are clearly defined.	Compiling an extensive corpus that encompasses each single word in a specific language is a challenging task. Domain-specific tasks necessitate a substantial amount of annotated training data.

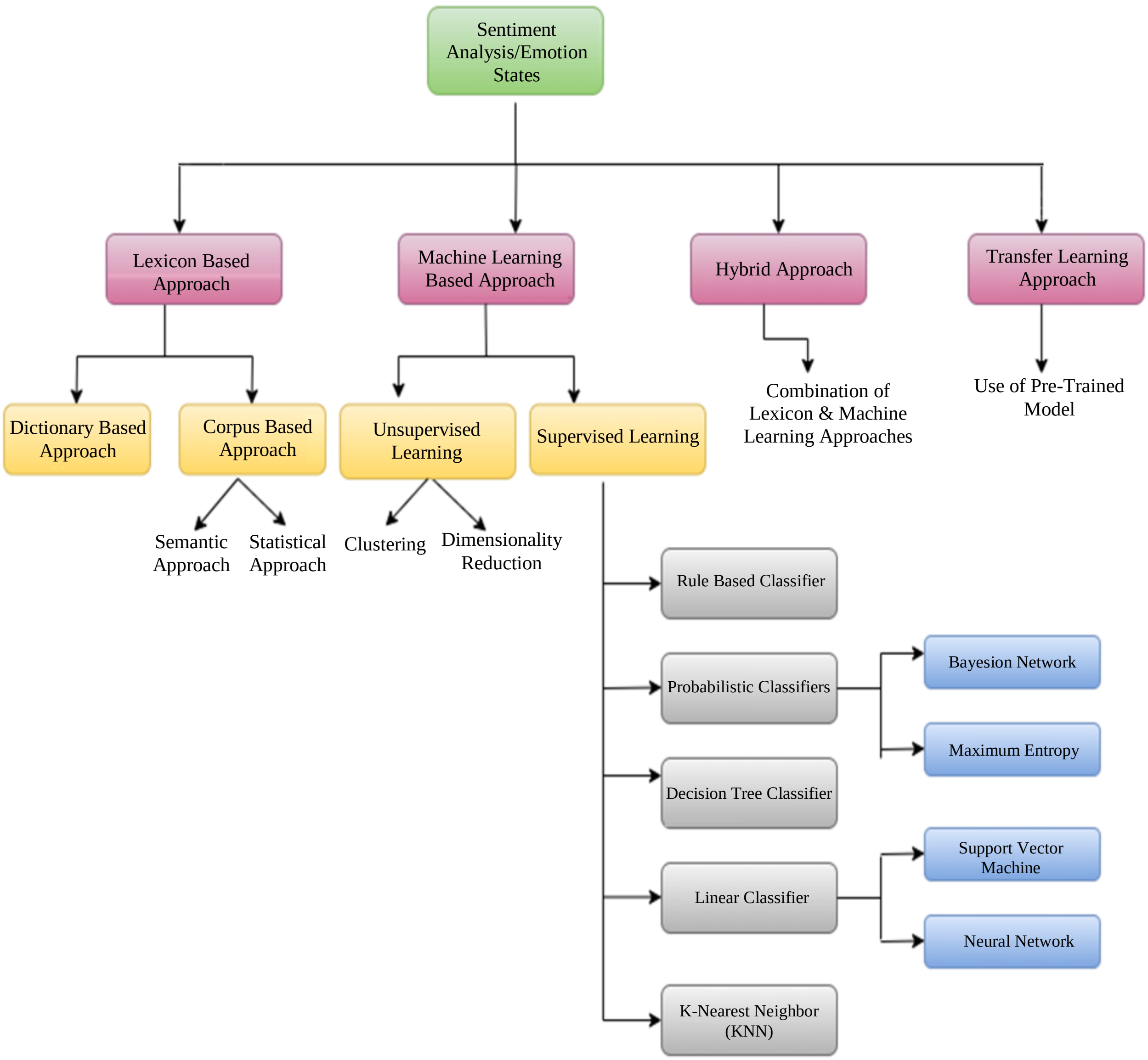

Fig. 7 Methodologies used for sentiment and emotion analysis

5.2. Machine Learning Methods

One approach to evaluating emotions using textual attributes is ML. Businesses use it to sift through user opinions on social media; it can work with or without human oversight. Socioeconomic systems and business outcomes have been profoundly affected by this. ML techniques, particularly supervised methods have enhanced the accuracy of SA. Training time and processing complexity are both lowered using unsupervised approaches.

Unsupervised SA seeks latent patterns in data through dictionaries, ontologies, databases, and information bases. Using three datasets-hotel reviews, IMDb, and "Polarity" dataset based on movies. A method based on unsupervised fuzzy logic was employed to analyze the sentiment of text reviews. The accuracy of the fuzzy technique is higher than that of nonfuzzy methods, but as its word scores are based on lexicons, it is vulnerable to first assumptions [60].

Soumya and Pramod classified nearly 3K Malayalam tweets into binary classes using feature vectors and ML techniques, with RF outperforming NB at 95.6% [49]. Since K-means clustering algorithms efficiently manage complexity in linear time, they are well-suited for SA. Accuracy can be impacted by uncertainties and anomalies, though. Consequently, supervised learning methods are frequently employed for SA since they produce more precise outcomes. Some popular algorithms used in supervised ML for detecting sentiments are:

SVM has a proven track record in text categorization due to their ability to be trained using a wide range of features. The correct functioning of SVM as SA is well acknowledged [61]. Using Distant Supervision and EFWS, the study investigates Twitter SA to circumvent computing constraints. A unique heuristic based on polarity assessments is introduced, and the method establishes a connection between the subjectivity inside tweets and the training samples.

Results from experiments demonstrate increased efficiency and accuracy (80% and 85%, respectively) when EFWS is used. The model can speed up training to twice the baseline pace. To further improve performance, the classification makes use of SVM and Particle Swarm Optimization [62]. There are many practical uses of SA in real-world scenarios, and one specific area where it can be applied is gathering comments from travel bloggers. Search engines like Google and Yahoo! make it easy for customers to find reviews pertaining to particular areas, and numerous reviews are available on different platforms.

However, the webpages displaying search engine results remain excessively intricate for human visual perception. A study analyzed travel blog opinions across seven US and European locales using a character-centered N-gram model, NB, and SVM. With 87% accuracy, SVM demonstrated superiority. Increasing the variety of client input is the goal of longitudinal analysis [10].

Due to its high accuracy with large datasets, Support Vector Machines (SVM) have become the predominant method for SA. Decision Trees (DT) are hierarchical structures representing potential solutions to problems. They are easier to implement than SVM because they do not require large datasets for training. Four text classifiers, specifically NB, J48, BFTree, and OneR, are utilized for SA [63]. NB applies conditional probability to assign words to their appropriate categories, while J48 employs decision trees to predict target phrases.

BFTree is a distinct classifier that uses a heuristic known as information gain to identify the most suitable node. A new method called OneR restricts DT to a single level of depth, leading to a solitary rule with minimal error rates. Only a few academics have utilized J48, BFTree, and OneR algorithms for sentiment prediction.

OneR has higher accuracy in terms of classification percentage; however, NB demonstrates a superior learning rate. When working with fewer datasets, J48 and OneR algorithms show better performance, with OneR achieving higher accuracy rates. The Random Forest (RF) algorithm is commonly used to address overfitting and improve accuracy, outperforming the use of DT.

One way to estimate probability distributions in natural language applications is using the Maximum Entropy (ME) classification approach. For regional languages such as Gujarati and Tamil, it is difficult to locate similar corpora. Data from non-parallel corpora can be automatically retrieved by using a number of tools [64].

In order to address this issue, it is possible to identify any language combination as having parallel sentences by utilizing the principles of ME and minimal training data. This may be achieved by the utilization of JMaxAlign [65], a Java-based maximum entropy tool that is capable of accepting two parallel corpora as input. Maximum entropy classifiers have the ability to produce meaningful results for almost any combination of languages, even when influenced by factors such as linguistic similarity and domain. Performing sentiment categorization with high accuracy in large and unpredictable data sets poses a substantial challenge for SA.

Emotion word extraction issues are suggested to be addressed by the PLSA model, which stands for ME Probabilistic Latent SA. It builds a training corpus by mining Wikipedia for basic terms. The model partitions datasets using the k-fold approach. Despite its low data requirement, the ME ML approach could struggle with datasets that are too diverse. Table 3 presents a concise overview of publications that have employed different ML approaches.

Table 3. Compilation of ML-based sentiment analysis techniques

Author	Summary	Inference
Dake and Gyimah 2023 [66]	Developed and implemented a number of ML classifiers to sift through student-submitted qualitative feedback, drawing insightful conclusions from students' free-form comments in the classroom.	An accuracy of 63.79% was achieved with SVM. This work narrows the algorithmic range and focuses on the SA of qualitative student feedback in educational settings, limiting the exploration of more effective models.
Qorib et al., 2023 [67]	SA on data collected from Twitter related to Covid-19. TF-IDF + Linear SVC were employed for classification.	Achieved an accuracy of 96.75% on manually collected datasets. However, experimental outcomes determine which model performs the best.
Chen et al., 2016 [68]	Utilized SVM and KNN for SA with TF-IGM feature extraction on newsgroup messages and economic news.	No specific results were provided.
Deng et al.,2014 [69]	Applied SVM for SA using term importance in documents and sentiment expression.	Achieved accuracies of 87% (Cornell), 8.70% (Amazon), and 88.00% (Stanford) on movie and product review datasets.
Balahur et al., 2012 [70]	Utilized SVM with lexicon-based feature extraction for SA on the emotional corpus.	Achieved precision of 62.89%, recall of 57.47%, and F-measure of 60.06%.
Kermani et al., 2020 [71]	The study uses an ML-based approach and genetic algorithm to calculate feature weight using Python software and Twitter datasets such as Stanford test corpus and STS-Gold.	The proposed TSA method exhibits improved efficiency, albeit with poor time complexity, which is particularly noticeable when handling large Twitter datasets.
Bibi et al., 2020 [72]	A study utilized clustering and classification techniques, including majority voting, k-means, SVM, and NB classifiers in WEKA, to analyze Twitter datasets like HCR, SS-Tweet, and STS-Test, focusing on SA in tweets.	Cooperative clustering using majority voting demonstrates superior quality compared to other methods.
Hassonah et al., 2020 [73]	Employed SVM with the combination of feature selection methods for SA on social tweets.	Successfully reduced the set of features by 97%.
Sehar et al., 2021 [74]	Utilized DL with BLSTM model for multimodal SA on Urdu language-based reviews.	Achieved accuracies of 84.32% (unimodal) and 95.35% (multimodal) on review videos from YouTube.
Belinda et al., 2022 [75]	Employed Multinomial NB with TF-IDF for SA on depression and anxiety datasets.	Achieved an accuracy of 96.15% on manually collected datasets.
Omuya et al., 2022 [76]	Utilized Naïve Bayes, SVM, and KNN algorithms with PCA for sentence-level SA on tweets.	Achieved accuracies of 98% (KNN), 90% (SVM), and 99% (NB) on the Sentiment140 dataset.
Muaad et al., 2022 [24]	Employed MNB, BNB, SGD, SVC, and LR algorithms with BoW and TF-IDF for sentence-level SA on the Arabic dataset.	Proposed technique tailored specifically for Arabic text.
Mohammed and Kora, 2022 [77]	Employed Ensemble DL for SA on general tweets with 6 Arabic and English corpus datasets.	Experimental research shows ensemble methods effectively reduce generalization errors and high variance in classifiers, but model complexity increases.
Lin Xiang, 2022 [78]	Utilized RF with improved TF-idf and chi-square for SA on literary texts.	Enhanced accuracy by 1.8% on average.

5.3. Hybrid Method

The term "hybrid" describes an approach to opinion mining that combines lexical and ML. SA employs a hybrid approach that combines statistical and knowledge-based methodologies to accurately identify polarity. The hybrid technique can yield improved results by surpassing the limitations of each conventional model. A novel approach, known as RFSVM, has been introduced, which combines the use of SVM and RF. Upon analyzing a dataset of 1000 reviews, it was determined that RFSVM achieved an accuracy of 82.4%, surpassing both SVM and RF. SVM achieved an accuracy level of 82.4%, while RF achieved 81% accuracy [79].

A model for summarizing customer reviews is being developed, which examines consumer behavior and decisions using techniques from NLP and Long Short-Term Memory (LSTM). As part of the model, pre-processing of data, extraction of features, and classification of sentiment was done. In order to classify sentiment, the model employs LSTM, which employs a hybrid approach that combines characteristics linked to reviews and aspects. According to experimental evaluations, the model achieves an average F1-score of 92.81%, a recall of 91.63%, and a precision of 94.46% [80]. Hybrid models outperform solo models, but there's room for improvement as models become complex during hybridization.

5.4. DL Approach

DL, a subfield of ML, has shown significant growth since its inception in 2006. DL is now being extensively studied due to its potential advantages compared to other approaches. One of the main reasons is that older methods, such as lexicon-based methods, require a hard and time-consuming process of manually crafting features. Moreover, individuals struggle to efficiently apply their knowledge to other disciplines or domains.

Furthermore, DL does not necessitate feature engineering as it automatically generates the features during the network training phase. It has the ability to process vast quantities of data beyond the limitations of typical ML approaches. More people use social media, especially Twitter, because of the coronavirus outbreak. In this study [81], SA, with the BERT model, learn how people feel about the illness and how eager they are to get it done. Global tweets and tweets made in India are the two datasets used for the analysis. The significance of physical separation in stopping the transmission of the virus is highlighted by the results, which reveal a validation accuracy of 94%. The basic RNN has been employed in various NLP tasks due to its ability to incorporate information from previous time steps to predict the current time step, effectively utilizing earlier data and functioning as a memory by retaining specific information in a sequence. The primary limitation of a normal Recurrent Neural Network (RNN) is its inability to retain long-term connections within a sequence, which is caused by the issues of vanishing and bursting gradient descent.

This issue can be overcome by employing different variations of Recurrent Neural Networks (RNN) such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional Long Short-Term Memory (Bi-LSTM). These RNN models are widely utilized in SA and NLP activities. Combining ontologies for semantic feature extraction, Word2vec for corpus conversion, and Convolutional Neural Network (CNN) for opinion mining, this research [82] proposes an effective method for SA. With an F-measure ranging from 86.03% to 88.52%, the method exceeds competing recall, accuracy, and precision methods. Particle swarm optimization is employed for parameter adjustment, and the ontology-based CNN implementation decreases false-positive and false-negative rates. Comparatively, the method performs better than other cutting-edge methods. Table 4 presents a concise overview of papers that have employed different DL methods.

Table 4. The collection of SA techniques based on DL

Author	Summary	Inference
Huang et al., 2022 [83]	Enhancement of the LSTM network "AEC-LSTM" by incorporating an attention mechanism and Emotional Intelligence (EI). Two benchmarked datasets, IMDB Yelp-14 and the other two, JDReview and SinaWeibo, were collected and used by web crawlers.	The current model's focus on aspect-level sentiment classification could be enhanced by incorporating diverse sentiment dictionaries for more nuanced analysis.
Aygun et al., 2023 [84]	A study using 928,402 tweets related to policy, health, media and others from English and Turkish users to find SA on six COVID-19 vaccines using 4 different versions of BERT.	The proposed model produced a representative sample of user views on vaccination with an accuracy rate of 87%. Further refinement is required to explore new NLP models to enhance accuracy and depth and expand SA to include additional countries and languages.
Alharbi et al., 2019 [85]	CNN with Word2vec embeddings on SemEval 2016 Twitter SA dataset	Attained an accuracy of 82.63% in SA using CNN with Word2vec embeddings.
Vohra and Garg., 2023 [86]	The purpose of this research is to analyze the tone of more than 450K manually collected tweets about remote employment using a CNN with FastText embeddings.	Achieved 92.6% accuracy. There is a positive tendency toward working from home, as shown by the results: 54.41% positive, 24.50% negative, and 21.09% neutral feelings.
Paramesha et al., 2023 [87]	Using a cross-domain movie review dataset, SA for multiclass polarity classification is done using a DL method like BERT.	An accuracy of 96.3% was achieved, but the work overlooks potential real-world implementation challenges, leaving uncertainties about its practical applicability and potential limitations in diverse environments.

Chandra and Kulkarni, 2022 [88]	Leveraging DL models like BERT, semantic and sentiment aspects of translations of selected chapters and verses of The Bhagavad Gita was done.	Despite differences in language and vocabulary, the semantic analysis showed that many translations had similar meanings and used comparable terms in terms of relevance. Expert Sanskrit translators should evaluate feelings in future projects.

5.5. Ensemble Based Approach

Combining numerous models into one stronger one is the goal of ensemble learning. Logistic Regression, SVM, NB, and RF were used in four datasets to analyze Twitter sentiment. It has been discovered that ensemble learning classifiers outperformed individual classifiers when using Bag-of-Words for feature extraction. By combining logistic regression with a stochastic gradient descent classifier, a voting classifier for SA was created. The vote classifier showed promise in sentiment recognition with 79% and 80% accuracy using TF and TF-IDF approaches, respectively [89]. The summarized study connected to the ensemble is shown in Table 5 in tabular format. The overall distribution of the learning algorithm for SA is depicted in Figure 8**.**

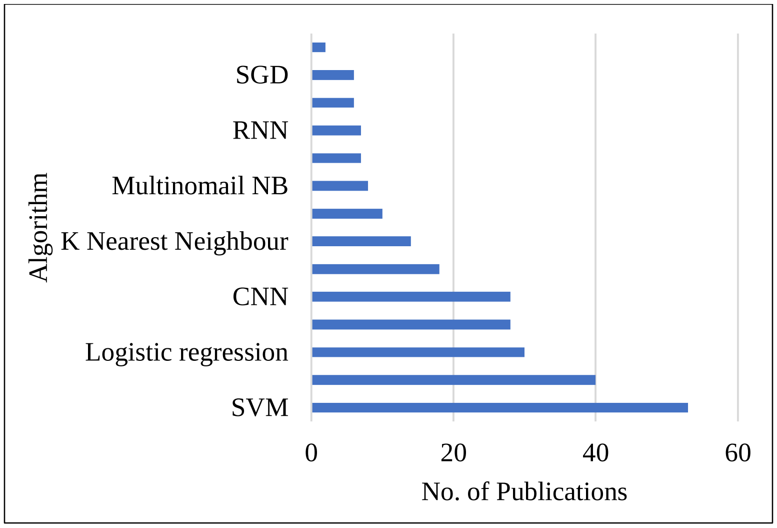

Fig. 8 Distribution based on mostly used learning algorithm for SA

Table 5. The SA techniques collection built on ensemble ML

Author	Summary	Inference
Gaye et al., 2020 [90]	Implemented LR-SGD with TF-IDF for SA on women's clothing reviews and hatred speech detection	Achieved 79% accuracy and 81% F1 score using the Voting Classifier (LR-SGD) with TF-IDF.
M Naz et al., 2019 [91]	The ensemble of KNN, NB, and SVM classifiers was done using FOA and mRMR feature selection techniques on the MATLAB tool. This research made use of a Twitter dataset that had been obtained from Blitzer's collection via the UCI repository.	Ensemble with used feature selection improves accuracy. The model's effectiveness can be further evaluated by testing it with real-time datasets.
Chiong et al.,2021 [92]	Gradient Boosting ensemble method was proposed to identify posts indicating depression on social media using 90 distinct characteristics. The method achieved an accuracy of 98%.	While ensemble models did the best overall, the potential utilization of SA datasets, readily available in larger sizes, for depression detection in social media texts will be explored.
Aslam et al.,2020 [93]	This work uses SA and emotion detection for cryptocurrency market value forecasting and the DL ensemble model combining LSTM and GRU features.	In order to improve prediction models, future studies should focus on combining SA with data from the Bitcoin market in real-time.
Aurangzeb et al.,2021 [94]	Using 7 datasets on multiple domains, an Evolutionary Ensembler (EEn) using SVM+GA based ensemble method has been proposed using BOW and word2vec.	Scaling of method required on a larger dataset.

Large Language Models (LLMs) are essential in natural language processing, focusing on tasks like text generation, machine translation, and chatbot interactions. They use extensive training data and advanced transformer architectures with over 100 billion parameters.

LLMs have significantly reshaped sentiment analysis research by utilizing deep learning and mass pre-learning on diverse textual data [7, 95]. They have demonstrated remarkable abilities in understanding context, creating coherent text, and capturing linguistic patterns. Studies have highlighted LLMs' effectiveness in capturing contextual information, understanding sarcasm, and dealing with complex linguistic structures.

6. Model Evaluation

In contemporary data science, a wide range of assessment metrics are commonly employed to evaluate the performance of models [96, 97].

Table 6 provides a concise overview of these metrics. The selection of one algorithm over another is determined by comparing the evaluation results of multiple algorithms, making this feature highly significant. The research findings demonstrate that accuracy, precision, recall, and F1-score remain widely adopted metrics within the ML field. Figure 9 displays the distribution that was determined using the assessment criteria.

Table 6. SA evaluation parameter

Parameter	Summary
Precision (P)	The percentage of expected positive samples that are actually classified properly is the measure of precision.
Accuracy	As a percentage, accuracy measures how many instances were accurately predicted out of all the examples.
Sensitivity	Sensitivity also known as Recall (R). The percentage of accurately detected positive samples relative to all positive samples is called recall.
F Measure	Also known as the F1 Score, which can be between one and zero. Its harmonic mean of precision and recall is useful when balancing between the two metrics and useful when data is imbalanced in nature.
Specificity	The polar opposite of sensitivity is specificity. An indicator of how well the negative class was predicted is the real negative rate.
Geometric-mean	It creates a unified measure that considers both sensitivity and specificity, making it easier to evaluate.
Area Under Cover (AUC)	AUC is calculated by laying down the P positive samples and Q negative samples on top of the ROC curve and axis of coordinates [96].
Time	Different temporal aspects are considered across various studies. Some examine computation time, time of execution, training duration, and testing duration [97].

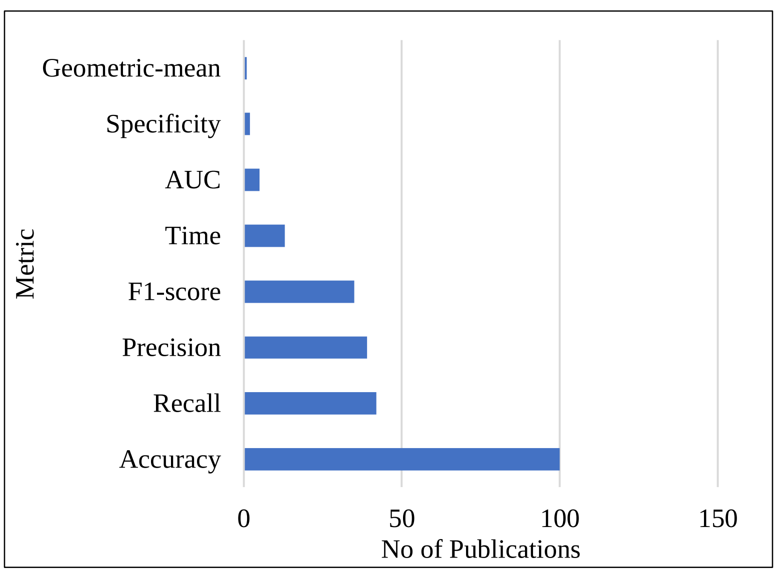

Fig. 9 Distribution determined by assessment criteria

7. Conclusion

Most research studies on sentiment classification are undertaken using data from the solitary language because there is a large number of corpora available in conventional dialects. The choice of SA methods is contingent upon several aspects, such as the specified task. Traditional ML methods are appropriate for subjective analysis, sentiment categorization and implicit linguistic recognition.

Lexicon is the most suitable tool for the aforementioned task, namely for extracting objects and aspects. Hybrid techniques do not effectively address the task of identifying sentiment spamming. DL is primarily employed for sentiment categorization problems involving large, non-textual, and intricate datasets. This survey shows that most past studies revolved around subjectivity detection while employing ML to discern emotional nuances only within textual content.

However, the abundance of multimedia data in modern times has sparked extensive discussions on emotion classification based on multimodality input, encouraging inquiries into SA. The utilization of ML and DL methods to assess sentiment in various forms of data, such as speech, image, video, etc., holds significant potential. Bilingual and multimodal SA offers extensive opportunities for enhancing accuracy by merging two or more modalities.

Additional focus is needed to improve the assessment of reviews expressed in languages other than English, including regional languages like Malayalam, Hindi, Bengali, Thai, Urdu, and more. This is because the current SA primarily focuses on applications for analyzing online reviews predominantly published in English. Currently, there is a lack of SA tools and software that are particular to certain domains and accessible to the general public.

However, academics may develop such tools in the future. Prior to making a purchase, users have the ability to conduct research on the opinions of previous buyers. The dataset has been the subject of only a handful of studies that evaluate melancholy detection, art, agriculture, and museums. The findings of this survey clearly indicate that researchers employ both supervised and unsupervised ML methods to assess SA. Traditional supervised ML methods like SVM and NB are widely used. Challenges and potential research areas are identified for interactive system advancement. It is suggested that future studies may explore the potential application of reinforcement learning techniques.

References

[1] Anton Borg, and Martin Boldt, “Using VADER Sentiment and SVM for Predicting Customer Response Sentiment,” *Expert Systems with Applications*, vol. 162, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Pansy Nandwani, and Rupali Verma, “A Review on Sentiment Analysis and Emotion Detection from Text,” *Social Network Analysis and Mining*, vol. 11, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Klaifer Garcia, and Lilian Berton, “Topic Detection and Sentiment Analysis in Twitter Content Related to COVID-19 from Brazil and the USA,” *Applied Soft Computing*, vol. 101, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Rui Ren, Desheng Dash Wu, and Tianxiang Liu, “Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine,” *IEEE Systems Journal*, vol. 13, no. 1, pp. 760-770, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] T. Tulasi Sasidhar, Premjith B., and Soman K.P., “Emotion Detection in Hinglish (Hindi+English) Code-Mixed Social Media Text,” *Procedia Computer Science*, vol. 171, pp. 1346-1352, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Alexander Ligthart, Cagatay Catal, and Bedir Tekinerdogan, “Systematic Reviews in Sentiment Analysis: A Tertiary Study,” *Artificial Intelligence Review*, vol. 54, pp. 4997-5053, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Jamin Rahman Jim et al., “Recent Advancements and Challenges of NLP-Based Sentiment Analysis: A State-of-the-Art Review,” *Natural Language Processing Journal*, vol. 6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, “Thumbs up? Sentiment Classification Using Machine Learning Techniques,” *Proceeding of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, pp. 79-86. 2002. [CrossRef] [Google Scholar] [Publisher Link]

[9] Peter D. Turney, “Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews,” *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417-424, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[10] Qiang Ye, Ziqiong Zhang, and Rob Law, “Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 6527-6535, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[11] Lal Khan et al., “Multi-Class Sentiment Analysis of Urdu Text Using Multilingual BERT,” *Scientific Reports*, vol. 12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Hewan Shrestha et al., “Natural Language Processing Based Sentimental Analysis of Hindi (SAH) Script an Optimization Approach,” *International Journal of Speech Technology*, vol. 23, pp. 757-766, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Akshi Kumar, and Teeja Mary Sebastian, “Sentiment Analysis: A Perspective on its Past, Present and Future,” *International Journal of Intelligent Systems and Applications*, vol. 4, no. 10, pp. 1-14, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[14] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni, "A Survey on Sentiment Analysis Methods, Applications, and Challenges," *Artificial Intelligence Review*, vol. 55, pp. 5731-5780, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] M. Devi Sri Nandhini, and G. Pradeep, "A Hybrid Co-Occurrence and Ranking-Based Approach for Detection of Implicit Aspects in Aspect-Based Sentiment Analysis," *SN Computer Science*, vol. 1, no. 3, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Mauro Dragoni, and Giulio Petrucci, "A Neural Word Embeddings Approach for Multi-Domain Sentiment Analysis," IEEE Transactions on Affective Computing, vol. 8, no. 4, pp. 457-470, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ashima Yadav, and Dinesh Kumar Vishwakarma, "Sentiment Analysis Using Deep Learning Architectures: A Review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335-4385, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ioannis Korkontzelos et al., "Analysis of the Effect of Sentiment Analysis on Extracting Adverse Drug Reactions from Tweets and Forum Posts," *Journal of Biomedical Informatics*, vol. 62, pp. 148-158, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[19] Andreas Lommatzsch et al., "Towards the Automatic Sentiment Analysis of German News and Forum Documents," *Innovations for Community Services*, pp. 18-33, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[20] Norah Saleh Alghamdi et al., "Predicting Depression Symptoms in an Arabic Psychological Forum," *IEEE Access*, vol. 8, pp. 57317-57334, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Zulfadzli Drus, and Haliyana Khalid, "Sentiment Analysis in Social Media and its Application: Systematic Literature Review," *Procedia Computer Science*, vol. 161, pp. 707-714, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Michelle Annett, and Grzegorz Kondrak, "A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs," *Advances in Artificial Intelligence*, pp. 25-35, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[23] Mondher Bouazizi, and Tomoaki Otsuki Ohtsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," *IEEE Access*, vol. 4, pp. 5477-5488, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[24] Abdullah Y. Muaad et al., "Arabic Document Classification: Performance Investigation of Preprocessing and Representation Techniques," *Mathematical Problems in Engineering*, vol. 2022, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Steven Bird, Ewan Klein, and Edward Loper, *Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit*, O'Reilly, 2009. [Google Scholar] [Publisher Link]

[26] P. Kathiravan, R. Saranya, and Sridurga Sekar, "Sentiment Analysis of COVID-19 Tweets Using TextBlob and Machine Learning Classifiers: An Evaluation to Show How COVID-19 Opinions Is Influencing Psychological Reactions of People's Behavior in Social Media," *Proceedings of International Conference on Data Science and Applications*, pp. 89-106, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Ted Kwartler, *Text Mining in Practice with R*, John Wiley & Sons, Ltd, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[28] Christopher Manning et al., "The Stanford CoreNLP natural Language Processing Toolkit," *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[29] Arfath Pasha et al., "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic," *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 1094-1101, 2014. [Google Scholar] [Publisher Link]

[30] Vaishali Kalra, and Rashmi Aggarwal, "Importance of Text Data Preprocessing & Implementation in RapidMiner," *Proceedings of the First International Conference on Information Technology and Knowledge Management*, vol. 14, pp. 71-75, 2017. [Google Scholar] [Publisher Link]

[31] Abdullah Ikram, Mohit Kumar, and Geetika Munjal, "Twitter Sentiment Analysis Using Machine Learning," *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, pp. 629-634, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] Ravinder Ahuja et al., "The Impact of Features Extraction on the Sentiment Analysis," *Procedia Computer Science*, vol. 152, pp. 341-348, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] Yousif A. Alhaj et al., "A Study of the Effects of Stemming Strategies on Arabic Document Classification," *IEEE Access*, vol. 7, pp. 32664-32671, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[34] Mohamed Elhag Mohamed Abo et al., "A Multi-Criteria Approach for Arabic Dialect Sentiment Analysis for Online Reviews: Exploiting Optimal Machine Learning Algorithm Selection," *Sustainability*, vol. 13, no. 18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[35] Despoina Antonakaki, Paraskevi Fragopoulou, and Sotiris Ioannidis, "A Survey of Twitter Research: Data Model, Graph Structure, Sentiment Analysis and Attacks," *Expert Systems with Applications*, vol. 164, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[36] Bhumika Gupta et al., "Study of Twitter Sentiment Analysis Using Machine Learning Algorithms on Python," *International Journal of Computer Applications*, vol. 165, no. 9, pp. 9-34, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[37] Nhan Cach Dang, María N. Moreno-García, and Fernando De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," *Electronics*, vol. 9, no. 3, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[38] Htet Htet Htun, Michael Biehl, and Nicolai Petkov, "Survey of Feature Selection and Extraction Techniques for Stock Market Prediction," *Financial Innovation*, vol. 9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[39] Mohammed Kasri, Marouane Birjali, and Abderrahim Beni-Hssane, "A Comparison of Features Extraction Methods for Arabic Sentiment Analysis," *Proceedings of the 4th International Conference on Big Data and Internet of Things*, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[40] Tomas Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," *arXiv*, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[41] Jeffrey Pennington, Richard Socher, and Christopher Manning, "GloVe: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532-1543, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[42] Abdullah Y. Muaad et al., "ArCAR: A Novel Deep Learning Computer-Aided Recognition for Character-Level Arabic Text Representation and Recognition," *Algorithms*, vol. 14, no. 7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[43] Siti Rohaidaha Ahmad, Azuraliza Abub Bakar, and Mohd Ridzwanb Yaakub, "A Review of Feature Selection Techniques in Sentiment Analysis," *Intelligent Data Analysis*, vol. 23, no. 1, pp. 159-189, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[44] Basant Agarwal, and Namita Mittal, "Optimal Feature Selection for Sentiment Analysis," *Computational Linguistics and Intelligent Text Processing*, pp. 13-24, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[45] Yousef Seyfari, and Akbar Meimandi, "A New Approach to Android Malware Detection Using Fuzzy Logic-Based Simulated Annealing and Feature Selection," *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 10525-10549, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[46] Abdur Rasool et al., "GAWA-A Feature Selection Method for Hybrid Sentiment Classification," *IEEE Access*, vol. 8, pp. 191850-191861, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[47] Ahmed Abbasi et al., "Selecting Attributes for Sentiment Classification Using Feature Relation Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 3, pp. 447-462, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[48] Hailong Zhang, Wenyan Gan, and Bo Jiang, "Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey," *2014 11th Web Information System and Application Conference*, Tianjin, China, pp. 262-265, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[49] Soumya S., and Pramod K.V., "Sentiment Analysis of Malayalam Tweets Using Machine Learning Techniques," *ICT Express*, vol. 6, no. 4, pp. 300-305, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[50] Gediminas Adomavicius, and YoungOk Kwon, "Improving Aggregate Recommendation Diversity Using Ranking Based Techniques," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 5, pp. 896-911, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[51] Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, and Rosa M. Calvo-Estévez, "A Wrapper Method for Feature Selection in Multiple Classes Datasets," *Bio-Inspired Systems: Computational and Ambient Intelligence*, pp. 456-463, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[52] Kang Leng Chiew et al., "A New Hybrid Ensemble Feature Selection Framework for Machine Learning Based Phishing Detection System," *Information Sciences*, vol. 484, pp. 153-166, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[53] Maryam Bahojb Imani, Mohammad Reza Keyvanpour, and Reza Azmi, "A Novel Embedded Feature Selection Method: A Comparative Study in the Application of Text Categorization," *Applied Artificial Intelligence*, vol. 27, no. 5, pp. 408-427, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[54] Rishabh Misra, and Prahal Arora, "Sarcasm Detection Using News Headlines Dataset," *AI Open*, vol. 4, pp. 13-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[55] Erik Cambria et al., "Affective Computing and Sentiment Analysis," *A Practical Guide to Sentiment Analysis*, pp 1-10, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[56] Ye Tian et al., "Facebook Sentiment: Reactions and Emojis," *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pp. 11-16, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[57] S.Ashika Parvin, M. Sumathi, and C. Mohan, "Challenges of Sentiment Analysis - A Survey," *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, pp. 781-786, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[58] Arindam Chatterjere et al., "Minority Positive Sampling for Switching Points-An Anecdote for the Code-Mixing Language Modeling," *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 6228-6236, 2020. [Google Scholar] [Publisher Link]

[59] Anna Jurek, Maurice D. Mulvenna, and Yaxin Bi, "Improved Lexicon-Based Sentiment Analysis for Social Media Analytics," *Security Informatics*, vol. 4, no. 1, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[60] Srishti Vashishtha, and Seba Susan, "Fuzzy Interpretation of Word Polarity Scores for Unsupervised Sentiment Analysis," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[61] Preeti Routray, Chinmaya Kumar Swain, and Smita Praya Mishra, "A Survey on Sentiment Analysis," *International Journal of Computer Applications*, vol. 76, no. 10, pp. 1-8, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[62] Vijay Gupta, and Punam Rattan, "Improving Twitter Sentiment Analysis Efficiency with SVM-PSO Classification and EFWS Heuristic," *Procedia Computer Science*, vol. 230, pp. 698-715, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[63] Jaspreet Singh, Gurvinder Singh, and Rajinder Singh, "Optimization of Sentiment Analysis Using Machine Learning Classifiers," *Human-Centric Computing and Information Sciences*, vol. 7, no. 1, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[64] Joseph Max Kaufmann, "JMaxAlign: A Maximum Entropy Parallel Sentence Alignment Tool," *Proceedings of COLING: Demonstration papers*, pp. 277-288, 2012. [Google Scholar] [Publisher Link]

[65] Xin Xie et al., "An Improved Algorithm for Sentiment Analysis Based on Maximum Entropy," *Soft Computing*, vol. 23, no. 2, pp. 599-611, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[66] Delali Kwasi Dake, and Esther Gyimah, "Using Sentiment Analysis to Evaluate Qualitative Students' Responses," *Education and Information Technologies*, vol. 28, no. 4, pp. 4629-4647, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[67] Miftahul Qorib et al., "COVID-19 Vaccine Hesitancy: Text Mining, Sentiment Analysis and Machine Learning on COVID-19 Vaccination Twitter Dataset," *Expert Systems with Applications*, vol. 212, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[68] Chen Kewen et al., "Turning from TF-IDF to TF-IGM for Term Weighting in Text Classification," Expert Systems with Applications, vol. 66, pp. 245-260, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[69] Zhi-Hong Deng, Kun-Hu Luo, and Hong-Liang Yu, "A Study of Supervised Term Weighting Scheme for Sentiment Analysis," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3506-3513, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[70] Alexandra Balahur, Jesús M. Hermida, and Andrés Montoyo, "Detecting Implicit Expressions of Emotion in Text: A Comparative Analysis," *Decision Support Systems*, vol. 53, no.4, pp. 742-753, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[71] Fatemeh Zarisfi Kermani, Faramarz Sadeghi, and Esfandiar Eslami, "Solving the Twitter Sentiment Analysis Problem Based on a Machine Learning-Based Approach," *Evolutionary Intelligence*, vol. 13, no. 3, pp. 381-398, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[72] Maryum Bibi et al., "A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis," *IEEE Access*, vol. 8, pp. 68580-68592, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[73] Mohammad A. Hassonah et al., "An Efficient Hybrid Filter and Evolutionary Wrapper Approach for Sentiment Analysis of Various Topics on Twitter," *Knowledge-Based Systems*, vol. 192, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[74] Urooba Sehar et al., "Urdu Sentiment Analysis via Multimodal Data Mining Based on Deep Learning Algorithms," *IEEE Access*, vol. 9, pp. 153072-153082, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[75] Carmel Mary Belinda M.J. et al., "Linguistic Analysis of Hindi-English Mixed Tweets for Depression Detection," *Journal of Mathematics*, vol. 2022, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[76] Erick Odhiambo Omuya, George Okeyo, and Michael Kimwele, "Sentiment Analysis on Social Media Tweets Using Dimensionality Reduction and Natural Language Processing," *Engineering Reports*, vol. 5, no. 3, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[77] Ammar Mohammed, and Rania Kora, "An Effective Ensemble Deep Learning Framework for Text Classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 8825-8837, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[78] Lin Xiang, "Application of an Improved TF-IDF Method in Literary Text Classification," *Advances in Multimedia*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[79] Yassine Al Amrani, Mohamed Lazaar, and Kamal Eddine El Kadiri, "Random Forest and Support Vector Machine Based Hybrid Approach to Sentiment Analysis," *Procedia Computer Science*, vol. 127, pp. 511-520, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[80] Gagandeep Kaur, and Amit Sharma, "A Deep Learning-Based Model Using Hybrid Feature Extraction Approach for Consumer Sentiment Analysis," *Journal of Big Data*, vol. 10, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[81] Mrityunjay Singh, Amit Kumar Jakhar, and Shivam Pandey, "Sentiment Analysis on the Impact of Coronavirus in Social Life Using the BERT Model," *Social Network Analysis and Mining*, vol. 11, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[82] Ravindra Kumar, Husanbir Singh Pannu, and Avleen Kaur Malhi, "Aspect-Based Sentiment Analysis Using Deep Networks and Stochastic Optimization," *Neural Computing and Applications*, vol. 32, pp. 3221-3235, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[83] Faliang Huang et al., "Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 9, pp. 4332-4345, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[84] İrfan Aygün, Buket Kaya, and Mehmet Kaya, "Aspect Based Twitter Sentiment Analysis on Vaccination and Vaccine Types in COVID-19 Pandemic with Deep Learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2360-2369, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[85] Ahmed Sulaiman M. Alharbi, and Elise de Doncker, "Twitter Sentiment Analysis with a Deep Neural Network: An Enhanced Approach Using User Behavioral Information," *Cognitive Systems Research*, vol. 54, pp. 50-61, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[86] Aarushi Vohra, and Ritu Garg, "Deep Learning Based Sentiment Analysis of Public Perception of Working from Home through Tweets," *Journal of Intelligent Information Systems*, vol. 60, no. 1, pp. 255-274, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[87] K. Paramesha et al., "Sentiment Analysis on Cross-Domain Textual Data Using Classical and Deep Learning Approaches," *Multimedia Tools and Applications*, vol. 82, no. 20, pp. 30759-30782, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[88] Rohitash Chandra, and Venkatesh Kulkarni, "Semantic and Sentiment Analysis of Selected Bhagavad Gita Translations Using BERT-Based Language Framework," *IEEE Access*, vol. 10, pp. 21291-21315, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[89] Anam Yousaf et al., "Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)," *IEEE Access*, vol. 9, pp. 6286-6295, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[90] Babacar Gaye, Dezheng Zhang, and Aziguli Wulamu, "A Tweet Sentiment Classification Approach Using a Hybrid Stacked Ensemble Technique," *Information*, vol. 12, no. 9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[91] Mehreen Naz, Kashif Zafar, and Ayesha Khan, "Ensemble Based Classification of Sentiments Using Forest Optimization Algorithm," vol. 4, no. 2, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[92] Raymond Chiong, Gregorious Satia Budhi, and Sandeep Dhakal, "Combining Sentiment Lexicons and Content-Based Features for Depression Detection," *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 99-105, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[93] Naila Aslam et al., "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," *IEEE Access*, vol. 10, pp. 39313-39324, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[94] Khursheed Aurangzeb, Nasir Ayub, and Musaed Alhussein, "Aspect Based Multi-Labeling Using SVM Based Ensembler," *IEEE Access*, vol. 9, pp. 26026-26040, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[95] Ogobuchi Daniel Okey et al., "Investigating ChatGPT and Cybersecurity: A Perspective on Topic Modeling and Sentiment Analysis," *Computers & Security*, vol. 135, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[96] Channabasava Chola et al., "Gender Identification and Classification of Drosophila Melanogaster Flies Using Machine Learning Techniques," *Computational and Mathematical Methods in Medicine*, vol. 2022, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[97] Nivin A. Helal et al., "A Contextual-Based Approach for Sarcasm Detection," *Scientific Reports*, vol. 14, 2024. [CrossRef] [Google Scholar] [Publisher Link]