



# Sentiment analysis: A survey on design framework, applications and future scopes

Monali Bordoloi<sup>1</sup> · Saroj Kumar Biswas<sup>2</sup>

Published online: 20 March 2023

© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

Sentiment analysis is a solution that enables the extraction of a summarized opinion or minute sentimental details regarding any topic or context from a voluminous source of data. Even though several research papers address various sentiment analysis methods, implementations, and algorithms, a paper that includes a thorough analysis of the process for developing an efficient sentiment analysis model is highly desirable. Various factors such as extraction of relevant sentimental words, proper classification of sentiments, dataset, data cleansing, etc. heavily influence the performance of a sentiment analysis model. This survey presents a systematic and in-depth knowledge of different techniques, algorithms, and other factors associated with designing an effective sentiment analysis model. The paper performs a critical assessment of different modules of a sentiment analysis framework while discussing various shortcomings associated with the existing methods or systems. The paper proposes potential multidisciplinary application areas of sentiment analysis based on the contents of data and provides prospective research directions.

**Keywords** Knowledge representation · Natural language processing · Sentiment analysis · Text analysis

## 1 Introduction

The advent of digitization accelerated the scope of the general public to express their sentiment or opinion on an online platform. An expert or general public nowadays desires to reach an optimal decision or opinion with the use of available opinionative data. Any online platform, such as an e-commercial website or a social media site, maintains a level of transparency, increasing its chance of influencing other users. However, a single topic

---

✉ Monali Bordoloi  
monali.bordoloi@gmail.com

Saroj Kumar Biswas  
bissarojkum@yahoo.com

<sup>1</sup> School of Computer Science and Engineering, VIT-AP University, Inavolu, Amaravati, Andhra Pradesh 522237, India

<sup>2</sup> Computer Science and Engineering Department, NIT Silchar, NIT Road, Silchar, Assam 788010, India

or item can possess millions of varied opinions on a single platform. The opinions or sentiments expressed can hold minute details or even a general opinion, which increases the research community's interest in further investigation. This was the beginning of the principle of sentiment analysis, also known as opinion mining. Sentiment analysis makes it easier to retrieve sentimental details, analyze opinionative/sentimental web data, and classify sentimental patterns in a variety of situations.

Sentiment analysis can be stated as the procedure to identify, recognize, and/or categorize the users' emotions or opinions for any service like movies, product issues, events, or any attribute as positive, negative, or neutral (Mehta and Pandya 2020). When sentiment is stated as a polarity in computational linguistics, it is typically treated as a classification task. When sentiment scores lying inside a particular range are used to express the emotion, the task is however regarded as a regression problem. Cortis et al. (2017) mentioned various research works where sentiment analysis is approached as either a classification or regression task. While analyzing the sentiments by assigning the instances sentiment scores within the range  $[-1, 1]$ , Cortis et al. (2017) discovered that there can be circumstances where the prediction is sometimes considered to be a classification task and other times to be regression. To solve the regression/classification problem, the authors developed a novel approach that combined the use of two evaluation methods to compute the similarity matrix. Therefore, mining and analysis of sentiment are either limited to positive/negative/neutral; or even deeper granular sentimental scale, depending on the necessity, topic, scenario, or application (Vakali et al. 2013).

In the last decade since the paper by Pang et al. (2002), a large number of techniques, methods, and enhancements have been proposed for the problem of sentiment analysis, in different tasks, at different levels. Numerous review papers on sentiment analysis are already available. It has been noted that the current studies do not give the scientific community a comprehensive picture of how to build a proper sentiment analysis model. A general, step-by-step framework that can be used as a guide by an expert or even by a new researcher would be ideal for designing a proper sentiment analysis model. Many of the existing surveys basically report the general approaches, methods, applications, and challenges available for sentiment analysis. The survey paper by Alessia et al. (2015) reports basic three levels of sentiment analysis, presents three types of sentiment classification approaches, discusses some of the available tools and methods, and points out four domains of applications of sentiment analysis. The study can be further extended to give more details about the different levels, methods/approaches, additional applications, and other related factors and areas. Wankhade et al. (2022) provided a detailed study of different sentiment analysis methods, four basic levels of sentiment analysis, applications based on domain and industries, and various challenges. The survey emphasizes several classification methods while discussing some of the necessary procedures in sentiment analysis. Instead of only concentrating on the procedures that are necessary for sentiment analysis, a detailed description of all the possible approaches is highly desirable as it can help in selecting the best among all for a certain type of sentiment analysis model. Each step/module of the sentiment analysis model should be discussed in detail to gain insight into which technique should be used given the domain, dataset availability, and other variables; or how to proceed further to achieve high performance. Further, applications of sentiment analysis are commonly described based on the domain or applicable industry. Possible application areas based purely on the dataset are rarely covered by recent review papers. Some of the survey papers focus on only one direction or angle of sentiment analysis. Multimodal sentiment analysis and its applications, as well as its prospects, challenges, and adjacent fields, were the main topics of the paper by Kaur and Kautish (2022). Schouten and Frasincar

(2015) focused on the semantically rich concept-centric aspect-level sentiment analysis and foreseen the rise of machine learning techniques in this context in the future. Verma (2022) addressed the application of sentiment analysis to build a smart society, based on public services. The author showed that understanding the future research directions and changes in sentiment analysis for smart society unfolds immense opportunities for related public services. Therefore, this survey paper aims to categorize sentiment analysis techniques in general, while critically evaluating and discussing various modules/steps associated with them.

This paper offers a broad foundation for creating a sentiment analysis model. Instead of focusing on specific areas, or enumerating the methodological steps in a scattered manner; this paper follows a systematic approach and provides an extensive discussion on different sentiment analysis levels, modules, techniques, algorithms, and other factors associated with designing an effective sentiment analysis model. The important contributions can be summarized as follows:

1. The paper outlines all the granularity levels at which sentiment analysis can be carried out, through appropriate representative examples.
2. The paper provides a generic step-by-step framework that can be followed while designing a simple as well as a high-quality sentiment analysis model. An overview of different techniques of data collection and standardization, along with pre-processing which significantly influences the efficiency of the model, are presented in this research work. Keyword extraction and sentiment classification having a great impact on a sentiment analysis model is thoroughly investigated.
3. Possible applications of sentiment analysis based on the available datasets are also presented in this paper.
4. The paper makes an effort to review the main research problems in recent articles in this field. To facilitate the future extension of studies on sentiment analysis, some of the research gaps along with possible solutions are also pointed out in this paper.

The remaining paper is organized into five different sections to provide a clear vision of the different angles associated with a sentiment analysis process. Section 2 provides knowledge of the background of sentiment analysis along with its different granularity levels. A detailed discussion of the framework for performing sentiment analysis is presented in the Sect. 3. Each module associated with designing an effective sentiment analysis is discussed in this section. Section 4 discusses different performance measures which can be used to evaluate a sentiment analysis model. Section 5 presents various possible applications of sentiment analysis based on the content of the data. Section 6 discusses the future scope of research on sentiment analysis. At last, Sect. 7 concludes the paper.

## 2 Background and granularity levels of sentiment analysis

The first ever paper that focused on public or expert opinion was published in 1940 by Stagner (1940). However, at that time studies were survey based. As reported in Mäntylä et al. (2018), the earliest computer-based sentiment analysis was proposed by Wiebe (1990) to detect subjective sentences from a narrative. The research on modern sentiment analysis accelerated in 2002 with the paper by Pang et al. (2002), where ratings on movie reviews were used to perform machine learning-based sentiment classification.

Pang et al. (2002) classified a document based on the overall sentiment, i.e., whether a review is positive or negative rather than based on the topic.

Current studies mostly concentrate on multilabel sentiment classification, while filtering out neutral opinions/sentiments. Due to the unavailability of proper knowledge of handling neutral opinion, the exclusion of neutral sentiment might lead to disruption in optimal decision-making or valuable information loss. Based on a consensus method, Valdivia et al. (2017) proposed two polarity aggregation models with neutrality proximity functions. Valdivia et al. (2018), filtered the neutral reviews using induced Ordered Weighted Averaging (OWA) operators based on fuzzy majority. Santos et al. (2020) demonstrated that the examination of neutral texts becomes more relevant and useful for comprehending and profiling particular frameworks when a specific polarity pre-dominates. Besides, there can be opinions that usually contain both positive and negative emotions as a result of noise. This kind of opinion is termed an ambivalence opinion, which is often misinterpreted as being neutral. Wang et al. (2020) presented a multi-level fine-scaled sentiment sensing and showed that the performance of the sentiment sensing improves with ambivalence handling. Wang et al. (2014) introduced the concept to classify a tweet with more positive than negative emotions into a positive category; and one with more negative emotions than the positive one into a negative sentiment category.

Computational linguistics, Natural Language Processing (NLP), text mining, and text analysis are different areas that are closely interlinked with the sentiment analysis process. The relationship between sentiment analysis and the different areas is summarized below:

### 1. *Computational linguistics*

Sentiment Analysis is a blend of linguistics and computer science (Taboada 2016; Hart 2013). Nowadays thousandths of human languages and other abbreviated or special languages exist, say the ones used in social media, which are used to convey thoughts, emotions, or opinions. People might use one single language or a combination of different languages, say for example Hinglish (a combination of Hindi and English) along with emoticons or some symbols to convey their messages. Computational linguistics assists in obtaining the computer-executable and understandable language from the vast source of raw languages through proper representation, to extract the associated sentiments properly. While developing formal theories of parsing and semantics along with statistical methods like deep learning, computational linguistics forms the foundation for performing sentiment analysis.

Linguistics knowledge aids in the development of the corpus set that will be used for sentiment analysis while understanding the characteristics of the data it operates on and determining which linguistic features may be applied. Data-driven or rule-based computer algorithms are designed to extract subjective information or to score polarity with the help of linguistic features, corpus linguistics, computational semantics, part of speech tagging, and the development of analytical systems for parsing. Connotations and associations are used to construct sentiment lexicons.

Recognition of sarcasm, mood classification, and polarity classification are some of the tasks covered by sentiment analysis, which is just a small subset of the discipline of computational linguistics. Approaches to classifying moods introduce a new dimension that is based on external psychological models. Methods for detecting sarcasm make use of ideas like “content” and “non-content” terms, which coexist in linguistic theory. Language models, such as Grice’s well-known maxims, are used to define sarcasm.

## 2. *Natural language processing*

NLP deciphers human language and makes it machine understandable. With the aid of NLP, the sentiments behind human-generated online comments, social media posts, blogs, and other information can be processed and represented by patterns and structures that can be used by software to comprehend and implement them. Sentiment analysis can be considered as a subset of NLP which helps users in opinionative/sentimental decision-making.

Different NLP tasks such as tokenization, stemming, lemmatization, negation detection, n-gram creation, and feature extraction aid in proper sentiment analysis. NLP-based pre-processing helps in improving the polarity classifier's performance by analyzing the sentiment lexicons that are associated with the subject (Chong et al. 2014). As a result, NLP facilitates text comprehension, accurately captures text polarity, and ultimately facilitates improved sentiment analysis (Rajput 2020; Solangi et al. 2018).

Advanced NLP techniques are often needed when dealing with emoticons, multilingual data, idioms, sarcasm, sense or tone, bias, negation, etc. Otherwise, the outcome can drastically deteriorate. If the NLTK's general stopwords list is utilized, words like not, nor, and no, for instance, are frequently deleted when removing stopwords during pre-processing. However, the removal of such words can alter the actual sentiment of the data. Thus, depending on its application, NLP tasks can either improve or deteriorate the result.

## 3. *Text mining*

Text messages, comments, reviews, and blog posts are excellent sources of sentimental information. The extraction of useful information and knowledge hidden in textual data is an important aspect of sentiment analysis. Mining the relevant information from textual data possesses multi-dimensional advantages such as improved decision-making, public influence, national security, health and safety, etc. (Zhang et al. 2021; Wakade et al. 2012). Text mining involves the use of statistical techniques to retrieve quantifiable data from unstructured text, and uses NLP to transform the unstructured text into normalized, structured data, which makes it suitable for sentiment analysis.

Sentiment analysis, however, is not just confined to text. In most cases, such as when a sarcastic comment is made, or while pointing a finger at someone and saying- "You are responsible!", the exact sentiment behind the plain text might not be conveyed properly. Non-text data like video, audio, and image are helpful in such a scenario to portray sentiment accurately.

## 4. *Text analysis*

A key part of sentiment analysis is extracting insightful information, trends, and patterns. To extract them from unstructured and semi-structured text data, text analysis is a process that supports sentiment analysis. Using techniques including word spotting, manual rule usage, text classification, topic modeling, and thematic analysis, the procedure helps in the extraction of meaning from the text. Text analysis can be used to specify individual lexical items (words or phrases) and observe the pattern.

Sentiment analysis, in contrast to basic text analytics, fundamentally shows the emotion concealed beneath the words, while text analytics analyses the grammar and relationships between words. Sentiment analysis essentially identifies whether a topic conveys a positive, negative, neutral, or any other sentiment; while text analysis is used to identify the most popular topics and prevalent ideas-based texts. In addition, it can be more challenging to specify the intended target in the context of sentiment conveyed, than it is to determine a document's general subject.

A textual document with numerous opinions would have a mixed polarity overall, as opposed to having no polarity at all (being objective). It is also important to distinguish the polarity and the strength of a conveyed sentiment. One may have strong feelings about a product being decent, average, or awful while having mild feelings about a product being excellent (due to the possibility that one just had it for a brief period before having an opinion.). Also, unlike topical (involving text) analysis, in many cases such as that of the quotes, it is critical to understand whether the sentiment conveyed in the document accurately reflects the author's true intentions or not.

Analyzing the existence of an important word in conjunction with the use of a sentiment score approach can help to uncover the most profound and specific insights that can be used to make the best decision in many situations. Areas of application for sentiment analysis aided by appropriate text analysis include strategic decision-making, product creation, marketing, competition intelligence, content suggestion, regulatory compliance, and semantic search.

## 2.1 Granularity levels

At present, a sentiment analysis model can be implemented at various granular levels according to the requirement and scope. There are mainly four levels of sentiment analysis that have gained a lot of popularity. They are document level (Pang et al. 2002; Li and Li 2013; Hu and Li 2011; Li and Wu 2010; Rui et al. 2013; Zhan et al. 2009; Yu et al. 2010), sentence or phrase level (Nguyen and Nguyen 2017; Wilson et al. 2005; Narayanan et al. 2009; Liu et al. 2013; Yu et al. 2013; Tan et al. 2012; Mullen and Collier 2004), word level (Nielsen 2011; Dang et al. 2009; Reyes and Rosso 2012; Bollegala et al. 2012; Thelwall and Buckley 2013; Li et al. 2014), and entity or aspect level (Li et al. 2012; Li and Lu 2017; Quan and Ren 2014; Cruz Mata et al. 2013; Mostafa 2013; Yan et al. 2015; Li et al. 2015a).

Some of the other research works concentrate on concept level (Zad et al. 2021; Tsai et al. 2013; Poria et al. 2013; Balahur et al. 2011; Cambria et al. 2022; Cambria 2013), link/user level (Rabelo et al. 2012; Bao et al. 2013; Tan et al. 2011), clause level (Kanayama and Nasukawa 2006; Liu et al. 2013), and sense level (Banea et al. 2014; Wiebe and Mihalcea 2006; Alfter et al. 2022) sentiment analysis. Some of the important levels of sentiment analysis are discussed in the following sub-sections. To understand the different levels, let us consider a customer review *R* as shown below.

*R* = "I feel the latest mobile from iPhone is really good. The camera has an outstanding resolution. It has a long battery life. I can even bear the mobile's heating problem. However, I feel it could have been a bit light weighted. Given the configurations, it is a bit expensive; but I must give a thumbs up for the processor."

In the following subsections, we will observe the analysis of review *R* based on different levels.

### 2.1.1 Document-level sentiment analysis

It aims to assess a document's emotional content. It assumes that the overall document expresses a single sentiment (Pang et al. 2002; Hu and Li 2011). The general approach of this level is to combine the polarities of each word/sentence in the document to find the overall polarity (Kharde and Sonawane 2016). According to document-level sentiment analysis, the overall sentiment of the document represented by review *R* is positive.

According to Turney (2002), there are two approaches to document sentiment classification namely term-counting and machine learning. Term counting measure derives a sentiment measure while calculating total positive and negative terms in the document. Machine learning approaches generally yield superior results as compared to term-counting approaches. In this approach, it is assumed that the document is focused on only one object and thus holds an opinion about that particular object only. Thus, if the document contains opinions about different objects, this approach is not suitable.

### 2.1.2 Sentence/phrase-level sentiment analysis

The sentiment associated with each sentence of a set of data is analyzed at this level of sentiment analysis. The general approach is to combine the sentiment orientation of each word in a sentence/phrase to compute the sentiment of the sentence/phrase (Kharde and Sonawane 2016). It attempts to classify a sentence as conveying either positive/negative/neutral/mixed sentiment or as a subjective or objective sentence (Katrekar and AVP 2005). Objective sentences are facts and do not convey any sentiment about an object or entity. They do not play any role in polarity determination and thus need to be filtered out (Kolkur et al. 2015). The polarity of a sentence in review R is found to be positive/negative/mixed irrespective of its overall polarity.

### 2.1.3 Word-level sentiment analysis

Through proper examination of the polarity of each and every word, this sentiment analysis level investigates how impactful individual words can be on the overall sentiment. The two methods of automatically assigning sentiment at this level are dictionary-based and corpus-based methods (Kharde and Sonawane 2016). According to Reyes and Rosso (2012), in corpus-based techniques, the co-occurrence patterns of words are used for sentiment determination. However, most of the time, statistical information needed for the determination of a word's sentiment orientation is large corpus dependent. The dictionary-based approaches use synonyms, antonyms, and hierarchies from lexical resources such as WordNet and SentiWordNet (SWN) to determine the sentiments of words (Kharde and Sonawane 2016). Such techniques assign positive, negative, and objective sentiment scores to each synset. If the words in review R such as outstanding, expensive, etc. are evaluated individually, different words within a particular sentence are observed to hold different polarities.

### 2.1.4 Aspect or entity-level sentiment analysis

For a specific target entity, this approach essentially identifies various aspects associated with it. Then, the sentiment expressed towards the target by each of its aspects is determined in this level of sentiment analysis. As a result, it can be divided into two different tasks, namely extraction of aspects and sentiment classification of aspects (Liu and Zhang 2012). For the different aspects such as resolution, weight, and price of the same product in review R, different sentiments are conveyed.



### 2.1.5 Concept-level sentiment analysis

Most of the time, merely using emotional words to determine sentiment or opinion is insufficient. To obtain the best results, a thorough examination of the underlying meaning of the concepts and their interactions is required. Concept-level sentiment analysis intends to convey the semantic and affective information associated with opinions, with the use of web ontologies or semantic networks (Cambria 2013). Rather than simply using word-cooccurrences or other dictionary-based approaches as in word-level sentiment analysis, or finding overall opinion about a single item as in document-level sentiment analysis; concept-level sentiment analysis generally makes use of feature spotting and polarity detection based on different concepts. E.g., For “long battery life” in review R is considered positive. However, a “long route” might not be preferable if someone wants to reach the destination in minimum time, and thus can be considered as negative. Tsai et al. (2013) made use of features of the concept itself as well as features of the neighboring concepts.

### 2.1.6 User-level sentiment analysis

User-level sentiment analysis takes into account the fact that if there is a strong connection among users of a social platform, then the opinion of one user can influence other users. Also, they may hold similar sentiments/opinions for a particular topic (Tan et al. 2011). At the user level, all the followers of the reviewer of review R may get influenced by this review.

### 2.1.7 Clause-level sentiment analysis

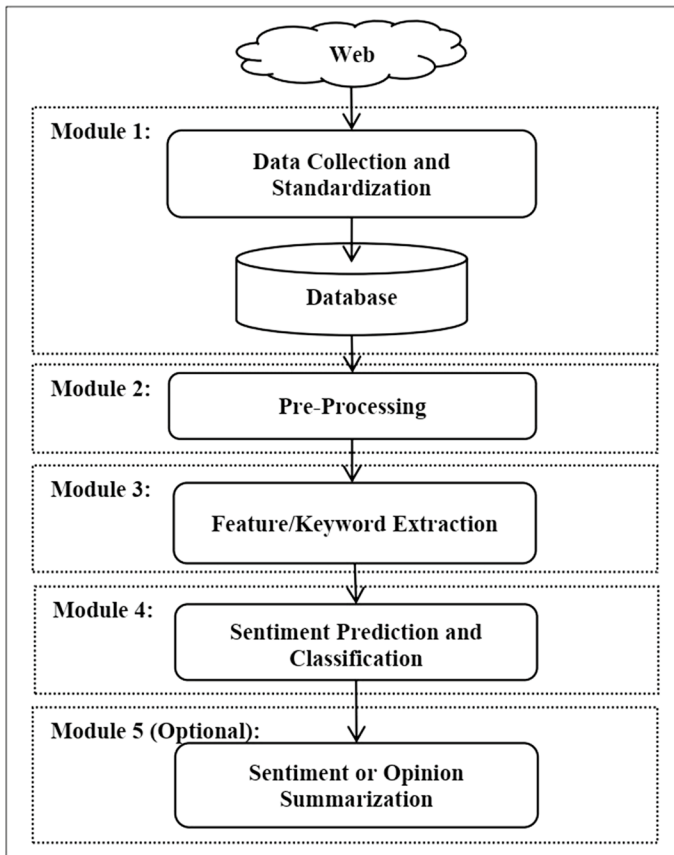
A sentence can be a combination of multiple clauses, each conveying different sentiments. The clauses in review R can be observed to represent opposing polarity because they are separated by the word “but”. Clause-level sentiment analysis focuses on the sentiment associated with each clause based on aspect, associated condition, domain, grammatical dependencies of the words in the clause, etc.

### 2.1.8 Sense-level sentiment analysis

The words which form a sentence can interpret different meanings based on their usage in the sentence. Specifically, when the same word has multiple meanings, the sense with which the word is used, can highly affect the sentiment orientation of the whole sentence or document. E.g., let us consider the word “bear” in review R. Is the word bear referring to the mammal bear? Otherwise, is it indicating the bearing (holding) of something? In what sense it is used? Is it used as a noun or a verb? In such a case, proper knowledge of the grammatical structure or word sense can contribute immensely to the determination of the appropriate sentiment of any natural language text. Thus, solving words’ syntactic ambiguity and performing word sense disambiguation (Wiebe and Mihalcea 2006) are vital parts of designing an advanced sentiment analysis model. Alfter et al. (2022) provided a sense-level annotated resource rather than word-level annotation and performed various experiments to explore the explanations of difficult words.

The analysis of the review R at different levels shows that the same review can have different interpretations based on the requirement. Single-level approaches work well in most cases. However, sometimes when the evaluation of sentiments is based on very short





**Fig. 1** General framework of sentiment analysis

document(s) or even very long document(s), the model may fail to handle the flexibility. To determine the polarity of the overall documents, Li et al. (2010) combined phrase-level and sentence-level sentiment analysis to design a multi-level model. Valakunde and Patwardhan (2013) advised following a ladder-like computation. In this technique, aspect or entity-level sentiment is employed to compute the sentence-level sentiments and then use the weightage of entities along with the sentence-level sentiments for evaluation of the complete document.

### 3 General framework of sentiment analysis

The evolution of sentiment analysis marks the emergence of different models by different experts. After going through more than 500 sentiment analysis models proposed till now, a general framework of sentiment analysis is presented in Fig. 1. The framework comprises mainly four modules along with an additional optional module. The modules perform collection and standardization of data; pre-processing of the dataset; extraction of features or keywords which represent the overall dataset; prediction or classification of the sentiments

associated with the keywords or the whole sentence or document according to the requirement; and summarization of the overall sentiment associated with the dataset. The different modules are discussed in detail below.

### 3.1 Data collection and standardization

With the growing platforms of expression, the type and format of expressing people's views, opinions, or sentiments on a particular subject is increasing. Among the different available types of data such as text, image, audio, or video, the research on textual data has gained momentum in the last few years. Currently, though multi-lingual text data has attracted few researchers, however, 90% of sentiment analysis studies, experimentation, and design concentrates mainly on English textual data.

The development, examination, and validation of a system typically depend on the quality and structure of data used for building, operating, and maintaining the model. The overall functionality of a model depends on the data used from the boundless and voluminous source of available data to a great extent. Many public data sources are available which are used by some researchers to design a sentiment analysis model. Publicly available dataset namely Blitzer's multi-domain sentiment data (Blitzer et al. 2007) is used by Dang et al. (2009). Public product reviews by Epinions (epinions.com) are also used by some of the researchers (Kharde and Sonawane 2016; Fahrni and Klenner 2008). UCI Machine Learning Repository provides standard datasets for sentiment namely Twitter data for Arabic Sentiment Analysis, Sentiment Labelled Sentences, Paper Reviews, Sentiment Analysis in Saudi Arabia about distance education during Covid-19, etc. The overwhelming rate of data production demands designing a system that keeps on updating the database from time to time to avoid generality or biased interest at a particular time. A manual approach to collecting a substantial volume of data is not a desirable practice. Thus, automatic big data collection techniques are indeed a vital aspect that must be keenly observed. Several tools or APIs have come up recently that help to collect data from online social or e-commercial platforms. Some of them are NodeXL, Google spreadsheet using Twitter Achiever, Zapier, Rapid Miner, Parsehub, BeautifulSoup in Python, WebHarvy, etc. Most of these tools or APIs help to collect real-time data. But the main problem occurs when someone desires to work with historical data; because many of these techniques such as Twitter API do not permit extracting tweets older than seven days. Building a standard database involves dealing with the unstructured information attached to the data from the internet. For a dataset representing a particular topic, proper standardization in an appropriate type, format, and context, extensively boosts the overall outcome of the analysis. To design a robust system, the homogeneity of the data must be maintained. Besides, proper labelling of the collected data can improve the performance of the sentiment analysis model. Different online labelling techniques are available nowadays. However, online labelling techniques are sometimes full of noise, which leads to lower accuracy of the system. Designing an automatic labelling system, which makes use of various statistical knowledge of the whole corpus and appropriate domain knowledge of words, proves to contribute more to enhancing the sentiment analysis process.

### 3.2 Pre-processing

The process of removing any sort of noise from a textual dataset and preparing a cleaned, relevant and well-structured dataset for the sentiment analysis process is called as

pre-processing. Appropriate pre-processing of any dataset noticeably improves the sentiment analysis process. For analyzing the sentiment of online movie reviews, a three-tier approach is adopted by Zin et al. (2017) to examine the effect of pre-processing task. In the first tier, they experimented with the removal of stopwords using the English stopwords list. The stopwords are the words such as the articles a, an, the, etc., which have no effective role in determining sentiment. In the second tier, the sentiment analysis is performed after the removal of stopwords and all other meaningless characters/words such as date (16/11/20), special characters (@, #), and words with no meaning (a+, a-, b+). In the third tier, more cleaning strategies are used, i.e., numbers and words having less than three characters are removed along with the stopwords and meaningless words. Their results demonstrate that the different combinations of the pre-processing steps show favorable improvement in the classification process; thus, establishing the significance of the removal of stopwords, meaningless words such as special characters, numbers, and words with less than three characters. Jianqiang (2015) found that replacing negations, and expanding acronyms have a positive effect on sentiment classification, however, the removal of URLs, numbers, and stopwords hardly changes the accuracy. Efficient pre-processing can increase the accuracy of a sentiment analysis model. To establish it, Haddi et al. (2013) combined various pre-processing methods using online reviews of movies and followed different steps such as cleaning online text, removal of white space, expansion of abbreviations, stemming, eliminating stopwords, and handling negation. Apart from these, they also considered feature selection as a pre-processing step. They used the chi-square method to filter out the less impactful features. To handle negation, a few researchers such as Pang et al. (2002), used the following words to tag the negation word until a punctuation mark occurs. However, authors of Haddi et al. (2013) and Dave et al. (2003) observed that the results before and after the tagging remain almost the same. Therefore, Haddi et al. (2013) reduced the number of tagged following words to three and two. Saif et al. (2014) observed that a list of pre-compiled stopwords negatively affects Twitter sentiment classification. However, with the use of pre-processing the original feature space is significantly reduced. Jianqiang and Xiaolin (2017) show that stopword removal, acronym expansion, and replacing negation are effective pre-processing steps. According to Jianqiang and Xiaolin, URLs and numbers do not contain useful information for sentiment analysis. They also found that reverting words with repeated characters shows fluctuating performance. This must be because, in some situations, a word such as goooooood gets replaced by goood. Thus, creating confusion about whether it should be interpreted as good or god. Such a situation may alter the actual polarity conveyed by the word. Therefore, reverting words with repeated characters is not recommendable.

### 3.3 Feature/keyword extraction

In a sentiment analysis model, the words and symbols within the corpus are mainly used as the features (O'Keefe and Koprinska 2009). Traditional topical text classification approaches are used in most sentiment analysis systems, in which a document is treated as a Bag of Words (BOW), projected as a feature vector, and then categorized using a proper classification technique. Experts use a variety of feature sets to boost sentiment classification efficiency, including higher-order n-grams (Pang et al. 2002; Dave et al. 2003; Joshi and Rosé 2009), word pairs and dependency relations (Dave et al. 2003; Joshi and Rosé 2009; Gamon 2004; Subrahmanian and Reforgiato 2008). Using different word-relation feature sets namely unigram (one word), bigram (two words), and dependency parsing,

Xia et al. (2011) performed sentiment classification using an ensemble framework. Wiebe and Mihalcea (2006) introduced a ground-breaking study focused on the Measure of Concern (MOC) to assess public issues using Twitter data and the most significant unigrams. While conducting text opinion mining, Sidorov et al. (2013) demonstrated the supremacy of unigrams, as well as other suitable settings such as minimal classes, the efficacy of balanced and unbalanced corpus, the usage of appropriate machine learning classifiers, and so on. Every word present in a dataset is not always important in the context of sentiment analysis. The difficulty of determining precise sentiment classifications has been increased by the continuous growth of knowledge. Even after cleaning the dataset with various pre-processing steps, using all of the data in the dataset can result in dimensionality issues, longer computation times, and the use of irrelevant or less significant features or terms. Especially in the case of higher dimensional and multivariate data, these problems become even worse. According to Li et al. (2017), a good word representation that captures sentiment is good at word sentiment analysis and sentence classification; and building document-level sentiment analysis dynamically based on words in need is the best practice. Keyword extraction is a method for extracting essential features/terms from textual data by defining particular terms, phrases, or words from a document to represent the document concisely (Benghuzzi and Elskeh 2020). If a text's keywords are extracted correctly, the text's subject can be thoroughly researched and evaluated, and a good decision can be made about the text. Given that, manually extracting keywords from such a large number of databases is a repetitive, time-consuming, and costly process, automated keyword extraction has become a popular field of research for most researchers in recent years. Automatic keyword extraction can be categorized into supervised, semi-supervised, and unsupervised methods (Beliga et al. 2015). The keywords are mainly represented using either Vector Space Model (VSM) or a Graph-Based Model (GBM) (Ravinuthala et al. 2016; Kwon et al. 2015). Once the datasets are represented using any of the VSM or GBM techniques, the keywords are extracted using simple statistics, linguistics, machine learning techniques, and hybridized methods (Bharti and Babu 2017). Simple methodologies that do not include training data and are independent of language and domain are included in the statistical keyword extraction methods. To identify keywords, researchers used frequency of terms, Term Frequency-Inverse Document Frequency (TF-IDF), co-occurrences of terms, n-gram statistics, PATricia (PAT) Tree, and other statistics from documents (Chen and Lin 2010). The linguistic approach examines the linguistic properties of words, sentences, and documents, with lexical, semantic, syntactic, and discourse analysis being the most frequently studied linguistic properties (HaCohen-Kerner 2003; Hulth 2003; Nguyen and Kan 2007). A machine learning technique takes into account supervised or unsupervised learning while extracting keywords. Supervised learning produces a system that is trained on a collection of relevant keywords followed by identification and analysis of keywords within unfamiliar texts (Medelyan and Witten 2006; Theng 2004; Zhang et al. 2006). All of these methods are combined in the hybrid method for keyword extraction. O'Keefe and Koprinska (2009) performed sentiment analysis using machine learning classifiers, which they validated using the movie review dataset. Along with the use of feature presence, feature frequency, and TF-IDF as feature weighting methods, they proposed SWN Word Score Groups (SWN-SG), SWN Word Polarity Groups (SWN-PG), and SWN Word Polarity Sums (SWN-PS) using words which are grouped by their SWN values. The authors suggest categorical Proportional Difference (PD), SWN Subjectivity Scores (SWNSS), and SWN Proportional Difference (SWNPD) as feature selection techniques. They discovered that feature weights based on unigrams, especially feature presence, outperformed SWN-based methods. Using different machine learning techniques; Tan and Zhang (2008) proposed

a model for sentiment analysis in three domains: education, film, and home, which was written in Chinese and used various feature selection techniques for the purpose. Mars and Gouider (2017) proposed a MapReduce-based algorithm for determining opinion polarity using features of consumer opinions and big data technologies combined with Text Mining (TM) and machine learning tools. Using a supervised approach, Kummer and Savoy (2012) suggested a KL score for providing weightage to features for sentiment and opinion mining. All these research works establish that the machine learning approach of keyword extraction when incorporated with any other techniques has a great scope in the field of sentiment analysis. There are different kinds of methods that are used to perform keyword extraction using VSM and GBM approaches. They are discussed in detail below.

### 3.3.1 Vector space model

In VSM, the documents are represented as vectors of the terms (Wang et al. 2015). VSM involves building a matrix  $V$  which is usually termed as a document-term matrix, where the rows represent the documents in the dataset, whereas columns correspond to the terms of the whole dataset. Thus, if the set of documents is represented by  $D = (d_1, d_2, \dots, d_m)$  and the set of terms/tokens representing the entire corpus is  $T = (t_1, t_2, \dots, t_n)$ , then the element  $dt_{ij} \in V_{m \times n}$ ,  $i = 1, 2, \dots, m$ , and  $j = 1, 2, \dots, n$  is assigned a weight  $w_{ij}$ . The weights can be assigned based on the word frequency associated with a document or the entire dataset. According to Abilhoa and De Castro (2014), the frequencies can be binary, absolute, relative, or weighted. Algorithms such as binary, Term Frequency (TF), TF-IDF, etc. are used in traditional term weighting schemes.

#### a. Binary

If document  $d_i$  contains the term  $t_j$ , the element  $dt_{ij}$  of a term vector is assigned a value 1 in the binary term weighting scheme, otherwise, the value 0 is assigned (Salton and Buckley 1988). It has the obvious drawback of being unable to recognize the most representative words in a text. Furthermore, using word frequency often helps to increase the importance of terms in documents.

#### b. TF

The limitation of the binary term weighting scheme motivates the use of term frequency as the weight of a term for a specific text. The number of times a word appears in a text is known as its term frequency. As a result, a value  $w_{ij}$  is assigned to  $dt_{ij}$  with  $w_{ij}$  equaling the number of times the word  $t_j$  appears in the document  $d_i$ . However, as opposed to words that appear infrequently in documents, terms that appear consistently in all documents have less distinguishing power to describe a document (Kim et al. 2022). This is an area where the TF algorithm falls short.

#### c. TF-IDF

The number of documents in the entire document corpus where a word appears is known as its document frequency. If a word has a higher document frequency, it has a lower distinguishing power, and vice versa. As a result, the Inverse Document Frequency (IDF) metric is used as a global weighting factor to highlight a term's ability to identify documents. Equation 1 (Zhang et al. 2020) may be used to describe a term's TF-IDF weight as follows:

$$W(t_k) = tf_k \cdot \log\left(\frac{m}{df_k}\right) \quad (1)$$

where,  $tf_k$  denotes the frequency of the term  $t_k$  in a specific document and  $df_k$  denotes the document frequency of the term  $t_k$ , i.e., the number of documents containing the term  $t_k$ . The total number of documents in the corpus is denoted by  $m$ .

Using the traditional term-weighting techniques, many experts tried to propose their improvised version. Some of them are TF-CHI (Sebastiani and Debole 2003), TF-RF (Lan et al. 2008), TF-Prob (Liu et al. 2009), TF-IDF-ICSD (Ren and Sohrab 2013), and TF-IGM (Chen et al. 2016).

### 3.3.2 Graph based model

A graph  $G$  is constructed in GBM, with each node or vertex  $V_i$  representing a document term or function  $t_i$  and the edges  $E_{ij}$  representing the relationship between them (Beliga et al. 2015). Nasar et al. (2019) showed that various properties of a graph, like centrality measures, node's co-occurrence, and others, play a significant role in keyword ranking. Semantic, syntactic, co-occurrence, and similarity relationships are some of the specific perspectives of graph-based text analysis. In GBM techniques, centrality measures tend to be the most significant deciding factor (Malliaros and Skianis 2015). The importance of a term is calculated by using the centrality measure, to calculate the importance of the node in the graph. Beliga (2014) presented the knowledge of nineteen different measures which are used for extraction purposes. Degree centrality, closeness centrality, betweenness centrality, selectivity centrality, eigenvector centrality, PageRank, TextRank, strength centrality, neighborhood size centrality, coreness centrality, clustering coefficient, and other centrality measures have been proposed so far. Some of the popular centrality measures are discussed below.

#### a. Degree centrality

Degree centrality is used to measure how often a term occurs with any other term. For a particular node, the total count of edges incident on it is used to measure the metric (Beliga 2014). The more edges that cross the node, the more significant it is in the graph. A node  $V_i$ 's degree centrality is measured using Eq. 2.

$$D_C(V_i) = \frac{|n(V_i)|}{|N| - 1} \quad (2)$$

where,  $D_C(V_i)$  represents node  $V_i$ 's degree centrality,  $|N|$  indicates the total count of nodes and  $|n(V_i)|$  represents the overall nodes linked with the node  $V_i$ .

#### b. Closeness centrality

Closeness centrality determines the closeness of a term with all other terms of the dataset. This metric calculates the average of the shortest distance from a given node to every other node in the graph. It is defined by Eq. 3 (Tamilselvam et al. 2017) as the reciprocal of the number of all node distances to any node, i.e. the inverse of farness.

$$C_C(V_i) = \frac{|N| - 1}{\sum_{V_j \in G} \text{dist}(V_i, V_j)} \approx \frac{N}{\sum_{V_j \in G} \text{dist}(V_i, V_j)}, \text{ if } (N \gg 1) \quad (3)$$

where,  $C_C(V_i)$  represents node  $V_i$ 's closeness centrality,  $|N|$  represents graph's node count, and  $\text{dist}(V_i, V_j)$  represents the shortest distance from node  $V_i$  to node  $V_j$ .

#### c. Betweenness centrality

This metric is used to see how often a word appears in the middle of another term. This metric indicates how many times a node serves as a bridge between two nodes on the shortest path. For a node  $V_i$ , it is calculated using Eq. 4 (Tamilselvam et al. 2017).

$$B_C(V_i) = \sum_{V_x \neq V_i \neq V_y \in G} \frac{\sigma_{V_x V_y}(V_i)}{\sigma_{V_x V_y}} \quad (4)$$

In Eq. 4,  $B_C(V_i)$  represents  $V_i$ 's betweenness centrality,  $\sigma_{V_x V_y}$  represents the overall shortest paths from node  $V_x$  to  $V_y$ , and the overall shortest paths from node  $V_x$  to  $V_y$  via  $V_i$  is represented by  $\sigma_{V_x V_y}(V_i)$ .

d. *Selectivity centrality*

Selectivity Centrality ( $S_C(V_i)$ ) (Beliga et al. 2015) is the average weight on a node's edges. As shown in Eq. 5,  $S_C(V_i)$  is equal to the fraction of strength of node  $s(V_i)$  to its degree  $d(V_i)$ .

$$S_C(V_i) = \frac{s(V_i)}{d(V_i)} \quad (5)$$

As shown in Eq. 6, node  $V_i$ 's strength,  $s(V_i)$ , is the summation of overall edge weights incident on ( $V_i$ ).

$$s(V_i) = \sum v_j E W_{V_i V_j} \quad (6)$$

b. *Eigenvector centrality*

This centrality measure determines the global importance of a term. It is calculated for a node using the centralities of the neighbors of the node. It is calculated using the adjacency matrix and a matrix calculation to determine the principal eigenvector (Golbeck 2013). Assume that  $A$  is a  $(n \times n)$  similarity matrix, with  $A = (\alpha_{V_i V_j})$ ,  $\alpha_{V_i V_j} = 1$  if  $V_i$  is bound to  $V_j$  and  $\alpha_{V_i V_j} = 0$ , otherwise. The  $i$ -th entry in the normalized eigenvector belonging to the largest eigenvalue of  $A$  is then used to describe the eigenvector centrality  $EV_C(V_i)$  of node  $V_i$ . Equation 7 (Bonacich 2007) shows the formula for eigenvector centrality.

$$\lambda EV_C(V_i) = \sum_{j=1}^N \alpha_{V_i V_j} EV_C(V_j) \quad (7)$$

where,  $\lambda$  is the largest eigenvalue of  $A$ . Castillo et al. (2015) suggested a supervised model with the use of degree and closeness centrality measures of a co-occurrence graph, to determine words belonging to each sentiment while representing existing relationships among document terms. Nagarajan et al. (2016) have also suggested an algorithm for the extraction of keywords based on centrality metrics of degree and closeness. For obtaining the optimal set of ranked keywords, Vega-Oliveros et al. (2019) used nine popular graph centralities for the determination of keywords and introduced a new multi-centrality metric. They found that all of the centrality measures have a strong relationship. The authors also discovered that degree centrality is the quickest and most efficient measure to compute. While experimenting with various centrality measures, Lahiri et al. (2014) also noticed that degree centrality makes keyword and key extraction much simpler. Abilhoa and De Castro (2014) suggest a keyword extraction model based on graph representation, and eccentricity and closeness



centrality measures. As a tiebreaker, they used the degree centrality. In several real-world models, disconnected graphs are common, and using eccentricity and closeness centralities to achieve the expected result often fails. Yadav et al. (2014) recommended extracting keywords using degree, eccentricity, closeness, and other centralities of the graph while emphasizing the semantics of the terms. With the use of Part of Speech (PoS) tagging, Bronselaer and Pasi (2013) presented a method to represent textual documents in a graph-based representation. Using various centralities, Beliga et al. (2015) proposed a node selectivity-driven keyword extraction approach. Kwon et al. (2015) suggested yet another ground-breaking keyword weighting and extraction method using graph. To improvise the traditional TextRank algorithm, Wang et al. (2018) used document frequency and Average Term Frequency (ATF) to calculate the node weight for extraction of keywords belonging to a particular domain. Bellaachia and Al-Dhelaan (2012) introduced the Node and Edge rank (NE-rank) algorithm for keyword extraction, which basically combines node weight (i.e., TF-IDF in this case) with TextRank. Khan et al. (2016) suggested Term-ranker, which is a re-ranking approach using graph for the extraction of single-word and multi-words using a statistical method. They identified classes of semantically related words while estimating term similarity using term embedding, and used graph refinement and centrality measures for extraction of top-ranked terms. For directed graphs, Ravinuthala et al. (2016) weighted the edges based on themes and examined their framework for keywords produced both automatically and manually. Using the PageRank algorithm, Devika and Subramaniaswamy (2021) extracted keywords based on the graph's semantics and centralities. The above studies show that centrality measures are a catalyst for effective sentiment analysis. This is because a powerful keyword's effect or position in determining the sentiment score is often greater than a weaker keyword. For the extraction of sentiment sentences, Shimada et al. (2009) suggested the use of a hierarchical acyclic-directed graph and similarity estimation. For sentences' sentiment representation, Wu et al. (2011) developed an integer linear programming-based structural learning system using graph. Using graphs, Duari and Bhatnagar (2019) also suggested keyword's score determination and extraction procedures based on the sentences' cooccurrence with a window size set to 2, position-dependent weights, contextual hierarchy, and connections based on semantics. In comparison to other existing models, their model has an excessively high dimensionality with terms in the text interpreted as nodes and edges representing node relationships in the graph. A variety of unsupervised graph-driven automated keyword extraction approaches is investigated by Mothe et al. (2018) using node ranking and varying word embedding and co-occurrence hybridization. Litvak et al. (2011) suggested DegExt, an unsupervised cross-lingual keyphrase extractor that makes use of syntactic representation of text using graphs. Order-relationship between terms represented by nodes is represented by the edges of such graphs. However, without a restriction on the maximum number of possible nodes which can be used, their algorithm generates exponentially larger graphs with larger datasets. As a result, dimensionality is one of the consequences of a graph-based keyword extraction procedure that must be regulated using appropriate means for sentiment analysis to be efficient. Chen et al. (2019) suggested extracting keywords using an unsupervised approach that relied solely on the article as a corpus. Words are ranked in their model based on their occurrence in strong motifs. Bougouin et al. (2013) assessed the relevance of a document's topic in order to suggest TopicRank, an unsupervised approach for extracting key phrases. However, it should be mentioned that their model does not have the optimal key selection approach. To retrieve topic-wise essential keywords,

Zhao et al. (2011) suggested a three-stage algorithm. Edge-weighting is used to rate the keywords (i.e., nodes) using two words' co-occurrence frequency, followed by generation as well as the ranking of candidate keyphrases. Shi et al. (2017) suggested an automated single document keyphrase extraction technique based on co-occurrence-based knowledge graphs, which learns hidden semantic associations between documents using Personalized PageRank (PPR). Thus, many experts have used co-occurrence graphs, as well as other graph properties such as centrality metrics, to demonstrate the effectiveness of these methods for keyword ranking in sentiment analysis.

### 3.4 Sentiment prediction and classification techniques

Different techniques have emerged till now for serving sentiment prediction and classification purposes. Several researchers group the techniques based on the applicability of the techniques, challenges, or simply the general topics of sentiment analysis. According to Cambria (2016), affective computing can be performed either by using knowledge-based techniques, statistical methods, or hybrid approaches. Knowledge-based techniques categorize text into affect categories with the use of popular sources of affect words or multi-word expressions, based on the presence of affect words such as 'happy', 'sad', 'angry' etc. Statistical methods make use of affectively annotated training corpus and determine the valence of affect keywords through word co-occurrence frequencies, the valence of other arbitrary keywords, etc. Hybrid approaches such as Sentic Computing (Cambria and Hussain 2015) make use of knowledge-driven linguistic patterns and statistical methods to infer polarity from the text.

Medhat et al. (2014) presented different classification techniques of sentiment analysis in a very refined and illustrative manner. Inspired by their paper, the current sentiment prediction and classification techniques are depicted in The evolution of sentiment analysis marks the emergence of different models by different experts. After going through more than 500 sentiment analysis models proposed till now, a general framework of sentiment analysis is presented in Fig. 2. The framework comprises mainly four modules along with an additional optional module. The modules perform collection and standardization of data; pre-processing of the dataset; extraction of features or keywords which represent the overall dataset; prediction or classification of the sentiments associated with the keywords or the whole sentence or document according to the requirement; and summarization of the overall sentiment associated with the dataset. The different modules are discussed in detail below. The techniques are examined thoroughly below, to assist in choosing the best sentiment analysis classification or prediction method for a particular task.

#### 3.4.1 Machine learning approach

The machine learning approach of sentiment classification uses well-known machine learning classifiers or algorithms along with linguistic features to classify the given set of data into appropriate sentiment classes (Cambria and Hussain 2015). Given a set of data, machine learning algorithms focus to build models which can learn from the representative data Patil et al. (2016). The extraction and selection of the best set of features to be used to detect sentiment are crucial to the models' performance Serrano-Guerrero et al. (2015). There are basically two types of machine learning techniques namely supervised and unsupervised. However, some researchers also use a hybrid approach by combining both these techniques.

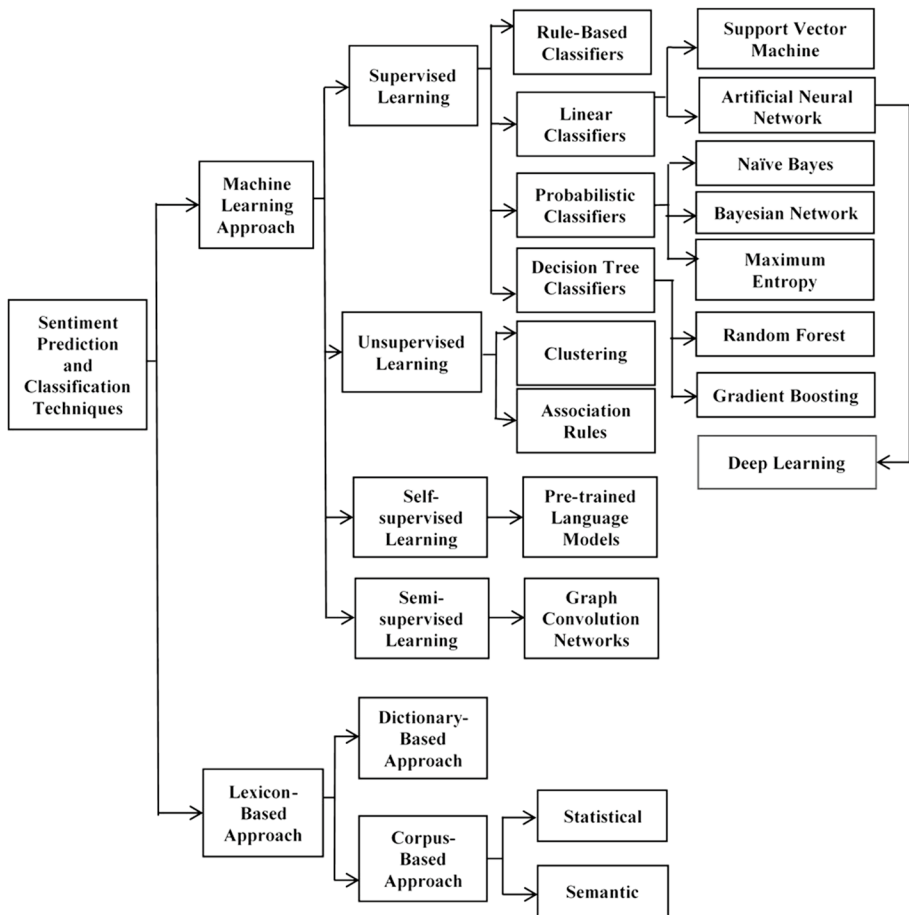


Fig. 2 Sentiment classification techniques

a. *Supervised learning*

The supervised machine learning approach is based on the usage of the initial set of labeled documents/opinions, to determine the associated sentiment or opinion of any test set or new document. Among the different supervised learning techniques Support Vector Machine (SVM), Naïve Bayes, Maximum Entropy, Artificial Neural Network (ANN), Random Forest, and Gradient Boosting are some of the most popular techniques which are employed in the sentiment analysis process. A brief introduction to each of these techniques is presented below; followed by a discussion on some of the research works using these algorithms either individually, in combination, or in comparison to each other.

- i. *Support vector machine* SVM is a kernel-based classifier that has gained popularity in different regression and classification problems. Many researchers established that Gaussian (or RBF) kernel function performs better for sentiment analysis (Kim et al. 2005; Li et al. 2015b). But, whenever a large number

of features is encountered, the RBF kernel does not provide suitable results. In the case of a very large dataset, the linear kernel function proves to be the best for text classification among all other different kernel functions used in the SVM classifier (Mullen and Collier 2004). The linear kernel function is represented as follows:

$$K(x_i, x_j) = x_i^T x_j \quad (8)$$

where,  $x_i$  and  $x_j$  are the input space vector and  $x_i^T$  is the transpose of  $x_i$ .

SVM classifier is basically designed for binary classification. However, if the model is extended to support multi-class classification, One-vs-Rest (OvR)/One against all or One-vs-One (OvO)/One against one strategy is applied for the SVM classifier (Hsu and Lin 2002). In OvR, the multi-class dataset is re-designed into multiple binary datasets, where data belonging to one class is considered positive while the rest are considered negative. Using the binary datasets, the classifier is then trained. The final decision on the assignment of a class is made by choosing the class which classifies the test data with the greatest margin. Another, strategy One-vs-One (OvO) can also be used, and thus choose the class which is selected by majority classifiers. OvO involves splitting the original dataset into datasets representing one class versus every other class one by one.

Ahmad et al. (2018) presented a systematic review of sentiment analysis using SVM. Based on the papers published during the span of 5 years, i.e., from 2012 to 2017, they found that a lot of research works are published either using SVM directly for analysis or in a hybrid manner or even for comparing their proposed model with SVM. Some of the recent studies that used SVM for sentiment analysis are listed in Table 1.

- ii. *Naïve Bayes* The probabilistic classifier Naïve Bayes classifies data based on the naive presumption that features are independent of one another. It is one of the simplest algorithms with low computational cost and relatively high accuracy. NB classifier uses the Bayesian method as shown in Eq. 9 to classify a document.

$$P(class_i | doc) = \frac{P(class_i)P(doc | class_i)}{P(doc)} \quad (9)$$

Given a document,  $P(class_i | doc)$  is the probability that a document belongs to a particular  $class_i$ .  $P(doc | class_i)$  is the probability that the document is present in that  $class_i$ ;  $P(class_i)$  and  $P(doc)$  are the probability of the  $class_i$  and the document respectively in the training set. In case of any other level of analysis such as sentence or word level sentiment analysis, simply doc in Eq. 9 is replaced with the required instance of the particular level.

There are basically two models which are commonly used for text analysis i.e., Multivariate Bernoulli Naïve Bayes (MBNB) and Multinomial Naïve Bayes (MNB) (Altheneyan and Menai 2014).

However, for continuous data, Gaussian Naïve Bayes is also used. MBNB is used for classification when multiple keywords (features) represent a dataset. In MBNB, the document-term matrix is built using BoW, where the keywords for a document are represented by 1 and 0 based on the occurrence or non-occurrence in the document.

**Table 1** Recent literature on sentiment analysis using SVM

Authors	Title of the paper	Contribution(s)
Hidayat et al. (2022)	Sentiment analysis of Twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as a classifier.	<ul style="list-style-type: none"><li>•Studied public opinion on Twitter regarding the development in Rinca Island using SVM and logistic regression.</li><li>•Used two types of Doc2Vec, distributed memory model of a paragraph vector (PV-DM) and a paragraph vector with a distributed bag of Words (PV-DBOW).</li><li>•The result of PV-DBOW with SVM, PV-DM with SVM showed the best results.</li></ul>
Cepeda and Jaiswal (2022)	Sentiment Analysis on Covid-19 Vaccinations in Ireland using Support Vector Machine	<ul style="list-style-type: none"><li>•Used tweets on the Covid-19 vaccination program in Ireland.</li><li>•A lexicon and rule-based VADER tool labeled the global dataset as negative, positive, and neutral. After that, Irish tweets were classified into different sentiments using SVM.</li><li>•Results show positive sentiment toward vaccines at the beginning of the vaccination drive, however, this sentiment gradually changed to negative in early 2021.</li></ul>
Mullen and Collier (2004)	Sentiment analysis using support vector machines with diverse information sources	<ul style="list-style-type: none"><li>•Uses SVMs to bring together diverse sources of potentially pertinent information, including several favourability measures for phrases and adjectives and, where available, knowledge of the topic of the text.</li><li>•Hybrid SVMs which combine unigram-style feature-based SVMs with those based on real-valued favourability measures obtain superior performance.</li></ul>
Zaimuddin and Selamat (2014)	Sentiment analysis using support vector machine	<ul style="list-style-type: none"><li>•The features were extracted using N-grams and different weighting schemes.</li><li>•Use of Chi-Square weight features to select informative features for the classification using SVM proves to improve the accuracy.</li></ul>

**Table 1** (continued)

Authors	Title of the paper	Contribution(s)
Luo et al. (2016)	Affective-feature-based sentiment analysis using SVM classifier	<ul style="list-style-type: none"> <li>• Considered text sentiment analysis as a binary classification.</li> <li>• The feature selection method of Chi-square Difference between the Positive and Negative Categories (CDPNC) was proposed to consider the entire corpus contribution of features and each category contribution of features.</li> <li>• The sentiment Vector Space Model (s-VSM) was used for text representation to solve data sparseness.</li> <li>• With the combination of document frequency with Chi-Square, the experimental results were found to be superior to other feature selection methods using SVM.</li> </ul>
Patil et al. (2014)	Sentiment analysis using support vector machine.	<ul style="list-style-type: none"> <li>• Stated that SVM acknowledges some properties of text like high dimensional feature space, few irrelevant features, sparse instance vector and also eliminates the need for feature selection with the ability to generalize high dimensional feature space.</li> <li>• The authors showed that the textual sentiment analysis performed better using SVM as compared to ANN.</li> </ul>
Prastyo et al. (2020)	Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized PolyKernel	<ul style="list-style-type: none"> <li>• The SVM analysis on the sentiments on general aspects using two-classes dataset achieved the highest performance in average accuracy, precision, recall, and f-measure.</li> <li>• Demonstrated that the SVM algorithm with the Normalized Poly Kernel can be used to predict sentiment on Twitter for new data quickly and accurately.</li> </ul>

Whenever the count of occurrence is considered, MNB is used. In MNB, the distribution is associated with vector parameters  $\theta_c = (\theta_{c1}, \theta_{c2}, \dots, \theta_{ci})$  for class  $c$ , where  $i$  is the number of keywords, and  $\theta_{ci}$  is the probability  $P(V_i | \text{Class}_c)$  of keyword  $V_i$  appearing in a dataset belonging to class  $c$ . For estimating  $\theta_c$ , a smoothed variant of maximum likelihood namely relative frequency counting is employed as shown below.

$$\hat{\theta}_{ck} = \frac{N_{ck} + \alpha}{N_c + \alpha n} \quad (10)$$

where,  $\alpha$  is the smoothing factor,  $N_{ci}$  is the number of times keyword  $k$  appears in the training set and  $N_c$  is the total number of keywords in class  $c$ .

To conduct a thorough investigation of the sentiment of micro-blog data, Le and Nguyen (2015) developed a sentiment analysis model using Naive Bayes and SVM, as well as information gain, unigram, bigram, and object-oriented feature extraction methods. Wawre and Deshmukh (2016) presented a system for sentiment classification that included comparisons of the common machine learning approaches Naive Bayes and SVM. Bhargav et al. (2019) used the Naive Bayes algorithm and NLP to analyze customer sentiments in various hotels.

- iii. *Maximum entropy* Maximum entropy classifier is a conditional probability model. Unlike the Naive Bayes classifier, it does not consider any prior assumptions such as the independence of keywords for the given set of data. Rather than using probabilities to set the parameters of the model, the maximum entropy classifier applies search techniques to determine the parameter set which maximizes the classifier's performance. After the determination of the document-term matrix, the training set is summarised in terms of its empirical probability distribution as shown in Eq. 11.

$$\tilde{P}(\text{doc}_i, c) = \frac{1}{N} * (\text{doc}_i, c) \quad (11)$$

where,  $N$  is the count of documents in the training set,  $n(\text{doc}_i, c)$  is the co-occurrence count of document  $\text{doc}_i$ , and the class  $c$  and  $\text{doc}_i$  comprises the contextual information of the document i.e. the sparse array.

Using the empirical probability distribution, maximum entropy models a given dataset by finding the highest entropy to satisfy the constraints of the prior knowledge. The unique distribution that shows maximum entropy is of the exponential form as shown in Eq. 12.

$$P^*(c | \text{doc}_i) = \frac{\exp(\sum_i \lambda_i f_i(\text{doc}_i, C))}{\sum_c \exp(\sum_i \lambda_i f_i(\text{doc}_i, C))} \quad (12)$$

Here,  $f_i(\text{doc}_i, C)$  is a keyword and  $\lambda_i$  is a parameter to be estimated. The denominator of Eq. 12 is a normalizing factor to ensure proper probability.

The flexibility offered by the maximum entropy classifier helps to augment syntactic, semantic, and pragmatic features with the stochastic rule systems. However, the computational resources and annotated training data required for the estimation of parameters for even the simplest maximum entropy model are very high. Thus, for large datasets, the model is not only expensive but is also



sensitive to round-off errors because of the sparsely distributed features. For the estimation of parameters, different methods such as gradient ascent, conjugate gradient, variable metric methods, Generalized Iterative Scaling, and Improved Iterative Scaling are available (Hemalatha et al. 2013). Yan and Huang (2015) used the maximum entropy classifier to perform Tibetan sentences' sentiment analysis, based on the probability difference between positive and negative outcomes. To identify the sentiment expressed by multilingual text, Boiy and Moens (2009) combined SVM, MNB, and maximum entropy describing different blogs, reviews, and forum texts using unigram feature vectors.

- iv. *Artificial Neural Network* ANN is a machine learning classifier that is designed based on the biological brain. In ANN, a set of fundamental processing units, known as a neuron, are connected and organized according to specific tasks. The network topology, weights between the neurons, activation function, bias, momentum, etc. together form the basis of learning in ANN. Among the different types of ANNs, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Recursive Neural Network (RecNN) are most commonly used for sentiment analysis. A convolutional layer is used by CNN to extract information from a larger piece of text. RNNs are particularly well suited to handle sequential data, such as text. They can be used in sentiment analysis to anticipate the sentiment when each token in a piece of text is processed. CNNs excel at extracting local and position-invariant features, but RNNs excel at a classification based on a long-range semantic dependency rather than local key phrases. As compared to shallow ANN which was initially introduced; deep ANN or generally known as the deep learning model has emerged as a powerful technique in the context of sentiment analysis.

*Deep learning (DL):* Deep Learning is essentially an ANN with three or more layers that has the capability to handle large datasets and their associated complexities such as non-linearity, intricate patterns, etc. It involves the transformation and extraction of features automatically, which facilitates self-learning as it goes by multiple hidden layers, in a way similar to humans. These advantages of deep learning lead to enhanced performance of a sentiment analysis model and thus have led to its popularity since 2015 for the same. The input features of many deep learning models are generally preferred to be word embeddings. Word embeddings can be learned from text data by using an embedding layer, Word2Vec, or Glove vectors. Word2Vec can be learned either by the Continuous Bag of Words (CBOW) or the Continuous Skip-Gram model. Some of the common deep learning algorithms include CNNs, RecNN, RNN, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Deep Belief Networks (DBN). The detailed study by Yadav and Vishwakarma (2020) on sentiment analysis using DL, has found that LSTM performs better than other popular DL algorithms.

Tembhurne and Diwan (2021) provided valuable insight into the usage of several architectural versions of sequential deep neural networks, such as RNN, for sentiment analysis of inputs in any form, including textual, visual, and multimodal inputs. Tang et al. (2015) introduced several deep NNs with the use of sentiment-specific word embeddings for performing word-level, sentence-level, and lexical-level sentiment analysis. To encode the sentiment polarity of sentences, the authors introduced different NNs including a prediction model and a ranking model. They discovered discriminative features from different domains

using sentiment embeddings to perform sentiment classification of reviews. According to the authors, the SEHyRank model shows the best performance among all the other proposed models. To fit CNN in aspect-based sentiment analysis, Wang et al. (2021) proposed an aspect mask to keep the important sentiment words and reduce the noisy ones. Their work made use of the position of aspects to perform aspect-based sentiment analysis in a unified framework. Hidayatullah et al. (2021) performed sentiment analysis using tweets on the Indonesian President Election 2019 using various deep neural network algorithms. According to the authors, Bidirectional LSTM (Bi-LSTM) showed better results as compared to CNN, LSTM, CNN-LSTM, GRU-LSTM, and other machine learning algorithms namely SVM, Logistic Regression (LR), and MNB. Soubraylu and Rajalakshmi (2021) proposed a hybrid convolutional bidirectional recurrent neural network, where the rich set of phrase-level features are extracted by the CNN layer and the chronological features are extracted by Bidirectional Gated Recurrent Unit (BGRU) through long-term dependency in a multi-layered sentence. Priyadarshini and Cotton (2021) suggested a sentiment analysis model using LSTM-CNN for a fully connected deep neural network and a grid search strategy for hyperparameter tuning optimization.

The Emotional Recurrent Unit (ERU) is an RNN, which contains a Generalized Neural Tensor Block (GNTB) and a Two-Channel Feature Extractor (TFE) designed to tackle conversational sentiment analysis. Generally, using ERU for sentiment analysis involves obtaining the context representation, incorporating the influence of the context information into an utterance, and extracting emotional features for classification. Li et al. (2022) employed ERU in a bidirectional manner to propose a Bidirectional Emotional Recurrent Unit (BiERU) to perform sentiment classification or regression. BiERU follows a two-step task instead of the three steps mentioned for simple ERUs. According to the source of context information, the authors presented two types of BiERUs namely, BiERU with global context (BiERU-gc) and BiERU with local context (BiERU-lc). As compared to c-LSTM (Poria et al. 2017), CMN (Hazarik et al. 2018), DialogueRNN (Majumder et al. 2019), and DialogueGCN (Ghosal et al. 2019), AGHMN (Jiao et al. 2020), BiERU showed better performance in most of the cases.

- v. *Random forest* Random Forest, similar to its name suggests, comprises a large number of individual choice trees that work as a troupe. Every individual tree in the random forest lets out a class expectation and the class with the most votes turn into our model's forecast. The major idea driving random forest is a straightforward however powerful one—the insight of groups. The explanation that the random forest model functions admirably is: A large number of relatively uncorrelated models (trees) working as a panel will beat any of the individual constituent models.

The low correlation between models is the key. Much the same as how speculations with low relationships meet up to shape a portfolio that is more prominent than the number of its parts, uncorrelated models can create group expectations that are more exact than any of the individual forecasts. The explanation behind this great impact is that the trees shield each other from their individual mistakes. While a few trees might not be right, numerous different trees will be correct, so as a gathering the trees can move the right way. So, the requirements for the random forest to perform well are:

There should be some real sign in our highlights so that models manufactured utilizing those highlights show improvement over random speculating.

The predictions made by the individual trees need to have low correlations with one another. As we realize that a forest is comprised of trees and more trees imply a more robust forest. Likewise, a random forest algorithm makes choice trees on information tests and afterward gets the forecast from every one of them, and lastly chooses the best arrangement by methods for casting a ballot. It is a gathering strategy that is superior to a solitary choice tree since it decreases the over-fitting by averaging the outcome.

Baid et al. (2017) analyzed the movie reviews using various techniques like Naïve Bayes, K-Nearest Neighbour, and Random Forest. The authors showed that Naïve Bayes performed better as compared to other algorithms. While performing sentiment analysis of real-time 2019 election twitter data, Hitesh et al. (2019) demonstrated that Word2Vec with Random Forest improves the accuracy of sentiment analysis significantly compared to traditional methods such as BoW and TF-IDF. This is because Word2Vec improves the quality of features by considering the contextual semantics of words.

- vi. *Gradient boosting* Boosting is one of the well-known ensemble techniques. Boosting builds the multiple models sequentially by assigning equal weights to each sample initially and then targets misclassified samples in subsequent models. Gradient Boosting Machine (GBM) is an ensemble technique that applies a decision tree as a base classifier. GBM constructs one tree at a time where each new tree helps to rectify errors caused by the earlier trained tree. Using a random forest classifier, the trees don't correlate with previously constructed trees, whereas, GBM relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. GBM can learn with different loss functions providing the ability to work efficiently with high-dimensional data. Moreover, sentiment analysis may encounter class imbalance issues, since the distribution of sentiments in real-world applications often displays issues of inequality. While handling the class imbalance problem using Synthetic Minority Oversampling Technique (SMOTE) algorithm and the Tomek links score, Athanasiou and Maragoudakis (2017) demonstrated that GBM outperforms decision trees, SVM, Naïve Bayes, and ANN. While using emoji and slang dictionaries, Prasad et al. (2017) found that GBM handles sarcasm in tweets much better than Random Forest, Decision Tree, Adaptive Boost, Logistic Regression, and Gaussian Naïve Bayes algorithm.

Jain and Dandannavar (2016) suggested a system for sentiment analysis of tweets based on an NLP-based technique and machine learning algorithms such as MNB and decision tree, which use features extracted based on various parameters. For sentiment analysis of online movie reviews, Sharma and Dey (2012) have developed a noteworthy comparison of seven current machine learning techniques in conjunction with various feature selection approaches. Tan and Zhang (2008) also introduced a similar work, in which sentiment analysis of various fields, such as education, movies, and houses, is carried out using various feature selection methods along with machine learning techniques. Depending on the applicability and need for better-quality models for sentiment analysis, experts in the field use a variety of cascaded and ensemble

approaches to combine machine learning algorithms with other existing options (Ji et al. 2015; Tripathy et al. 2015; Xia et al. 2011; Ye et al. 2009).

b. *Unsupervised learning*

In unsupervised learning, the models are trained using unlabeled datasets. This technique in most cases relies on clustering methods such as k-means clustering, expectation-maximization, and cobweb. Darena et al. (2012) used k-means clustering through the use of Cluto 2.1.2 to determine the sentiment associated with customer reviews.

c. *Self-supervised learning*

In self-supervised learning, the model begins with unlabeled datasets and then trains itself to learn a part of the input by leveraging the underlying structure of the data. Although the use of an unlabeled dataset gives this learning technique the notion of being unsupervised, they are basically designed to execute downstream tasks that are traditionally addressed by supervised learning. One of the self-supervised learning techniques which have gained a lot of popularity in recent years is the Pretrained Language Model (PML).

- i. *Pretrained language model (PLM)* PLMs are models that are built and trained previously, specifically using a large language dataset. In the case of sentiment analysis, the use of transfer learning by a pretraining model to add knowledge from the open domain to a downstream task to improve that task is extremely beneficial when limited language resources are available. The learning process of PLM involves processes such as missing or masked word prediction, next sentence prediction, sentence-order prediction, corrupted text reconstruction, autoregressive language modeling, etc. which aids in the sentiment prediction process.

Typical steps in the process of creating a sentiment analysis model from scratch usually involve making use of standard sentiment lexicons, sentiment scoring and data labeling by human experts, and proper parameter tuning of the model that performs well on the rest of the dataset. This procedure could be expensive and time-consuming. PLM makes it simpler for developers of sentiment analysis models to implement the model in less training time with improved accuracy, by providing extensive semantic and syntactic information with the usage of a few lines of code. PLM acts as a reusable NLP model for various tasks associated with sentiment analysis such as PoS tagging, lemmatization, dependency parsing, tokenization, etc. Thus, PLMs can be proved to be advantageous to solve similar new tasks using old experience, without training the sentiment analysis model from the scratch.

Chan et al. (2022) provided a detailed study on the evolution and advancement of sentiment analysis using pretrained models. Additionally, the authors covered various tasks of sentiment analysis, for which the pretrained models can be used. The early works on PML involved transferring a single pretrained embedding layer to the task-oriented network architecture. To cope with numerous challenges such as word sense, polysemy, grammatical structure, semantics, and anaphora, models are presently being improved to a higher representation level.

Bidirectional Encoder Representations from Transformers BERT (Devlin et al. 2018), NLTK (Loper and Bird 2002), Stanford NLP (Manning et al. 2014), Universal Language Model Fine-tuning (ULMFit) (Howard and Ruder 2018), Embeddings from Language Models (ELMo) (Sarzynska-Wawer et al. 2021) are some of the well-known PLMs that serve as open-source NLP libraries for sentiment analysis.

The performance of BERT was determined to be superior by Mathew and Bindu (2020) who thoroughly analyzed numerous PLMs that are frequently used for sentiment analysis.

Many pre-trained models use self-supervision strategies to learn the semantic content; however, give less importance to the sentiment-specific knowledge during the pre-training phase. There might also be a risk of overfitting associated with a pretraining model, which may lead to domain-specific sentiment mismatch between the source and the target domain. While dealing with social media related content, the PLM might cause biases in the results. The language in which the PLM was trained might differ from the language which is generally used in social media platforms. Further in-depth analysis and model development may be constrained if PLM behaves in a black-box manner. In a few cases, the PLM might not be able to handle the multi-class problem, if it was originally designed for identifying single or binary classes. This might also lead to ignorance/mishandling of one of the important classes, say neutral class, if the PLM was initially designed for handling positive and negative classes. Thus, while choosing a particular PLM model, we must consider the domain and data it was originally designed for. Also, a human expert might be required to validate the results, whenever required, to assure the quality of the sentiment analysis model.

Mao et al. (2022) provided an in-depth analysis of how PLMs are biased toward prompt-based sentiment analysis and emotion detection. According to the authors, the number of label classes, emotional label-word selections, prompt templates and positions, and the word forms of emotion lexicons leads to biased results. To address the issue of cross-domain tasks, Zhou et al. (2020) proposed SENTIX, a sentiment-aware model that learns the domain invariant sentiment knowledge during the pre-training phase. For addressing several factors related to sentiment analysis, experts have till now presented a variety of improvised modifications of the original PLMs. Some of them include Dynamic Re-weighting BERT (DR-BERT) (Zhang et al. 2022), BERT-based Dilated CNN (BERT-DCNN) (Jain et al. 2022), Attention-based ELMo (A-ELMo) (Huang and Zhao 2022), Contextual Sentiment Embeddings (CoSE) (Wang et al. 2022a), Extended Universal Language Model Fine-Tuning (Ext-ULMFiT) and Fine-Tuned (FiT-BERT) (Fazlourrahman et al. 2022), etc.

d. *Other learning techniques*

Many researchers combine supervised and unsupervised techniques to generate hybrid approaches or even semi-supervised techniques which can be used to classify sentiments (König and Brill 2006; Kim and Lee 2014). With new information generated every millisecond, finding a fully labeled large dataset representing all the required information is nearly impossible. In such a scenario, semi-supervised algorithms train an initial model on a few labeled samples and then iteratively apply it to the greater number of unlabelled data and make predictions on new data. Among various semi-supervised techniques, Graph

Convolution Network (GCN) (Kipf and Welling 2016; Keramatfar et al. 2022; Dai et al. 2022; Zhao et al. 2022; Lu et al. 2022; Yu and Zhang 2022; Ma et al. 2022) has recently gained the attention of researchers for performing sentiment analysis.

i. *Graph Convolution Network*

GCN is based on CNN which operates directly on graphs while taking advantage of the syntactic structure and word dependency relation to correctly analyze sentiment. GCNs learn the features by inspecting neighboring nodes. By using a syntactic dependency tree, a GCN model captures the relation among different words and links specific aspects to syntax-related words. Each layer of the multi-layer GCN architecture encodes and updates the representation of the graph's nodes using features from those nodes' closest neighbors. GCNs assist in performing node-level, edge-level, and graph-level prediction tasks for sentiment analysis, such as determining how connections on a social media platform affect the opinions of the users within that network, creating user recommendations based on connections between various products previously purchased, suggesting movies, etc. Generally, GCNs focus on learning the dependency information from contextual words to aspect words based on the dependency tree of the sentence. As a result, GCN has mainly attracted researchers in the field of aspect-based sentiment analysis.

Lu et al. (2021) built a GCN on the sentence dependency tree to fully utilize the syntactical and semantic information. Their methodology fixed the issues of incorrectly detecting irrelevant contextual words as clues for evaluating aspect sentiment, disregarding syntactical constraints, and long-range sentiment dependencies, which were present in earlier models. SenticGCN was proposed by Liang et al. (2022) to capture the affective dependencies of the sentences according to the specific aspects. To combine the affective knowledge between aspects and opinion words, the model performs aspect-based sentiment analysis using SenticNet along with GCN.

Along with the local structure information of a given sentence, such as locality, sequential knowledge, or syntactical dependency constraints within the sentence, global dependency information also holds importance in determining the sentiments accurately. Zhu et al. (2021) proposed a model named Global and Local Dependency Guided Graph Convolutional Networks (GL-GCN), where word global semantic dependency relations were revealed with the use of a word-document graph representing the entire corpus. An attention mechanism was adopted by the authors to combine both local and global dependency structure signals.

In general, the layers in GCN models are not devised distinctively for processing the aspect. To handle this issue, Chen et al. (2021) integrated GCN and co-attention networks for aspect-based sentiment analysis, to extract relevant information from contexts and remove the noise while considering colloquial texts. Tian et al. (2021) addressed the issues of the inability to learn from different layers of GCN, not considering dependency types, and lacking mechanisms for differentiating between various relations in the context of sentiment analysis. The authors utilized dependency types for aspect-based sentiment analysis with Type-aware GCN (T-GCN).

ii. *Lexicon-based approach*

Opinion terms are used in a lexicon-based approach to execute sentiment classification tasks. This method suggests that a sentence's or document's cumulative polarity is the sum of the polarities of individual terms or phrases (Devika et al. 2016). According to Zhang et al. (2014), in opinion lexicon methods, evaluated and

tagged sentiment-related words are counted and weighted with the use of a lexicon to perform sentiment analysis. This approach is based on sentiment lexicons, which are a compilation of recognized and pre-compiled terms, phrases, and idioms formed for traditional communication genres, according to Kharde and Sonawane (2016). More complex systems, such as dictionaries or ontologies, may also be used for this approach (Kontopoulos et al. 2013). Some of the recent lexicons available for sentiment analysis are discussed below in Table 2.

The lexicon-based approach is categorized into three methods: manual, dictionary-based, and corpus-based methods based on the various approaches to classification (Zhang et al. 2014). Because of the considerable time investment, researchers seldom use the manual approach, though it is often paired with the other two automated approaches.

a. *Dictionary-based approach*

Dictionary-based approach starts with a series of manually annotated opinion seed terms. The collection is then extended by searching through a dictionary such as WordNet (Miller et al. 1990) to find synonyms and antonyms. SWN (Baccianella et al. 2010) is one of the earliest thesauri and makes use of WordNet to assign positive, negative, and objective ratings to terms. The new words are added to the initial list after they have been discovered. The next iteration begins and the method continues until no new words need to be added after a particular point. While considering valence shifters (intensifiers, downtoners, negation, and irrealis markers), Read and Carroll (2009) proposed a word-level sentiment analysis model called Semantic Orientation CALculator (SO-CAL). In SO-CAL, lexicon-based sentiment classification is performed using dictionaries of sentiment-bearing terms annotated with their polarities and strengths.

The use of a dictionary for sentiment analysis suffers from one major drawback. This methodology does not adequately handle the domain and context-sensitive orientations of opinion terms.

b. *Corpus-based approach*

The corpus-based approach uses syntactic patterns or co-occurring patterns in a vast corpus to extend the original seed list of opinion terms (Cambria and Hussain 2015). It is very tough to generate a huge corpus using the corpus-based approach, to cover each and every English word. However, using a domain corpus has the advantage of allowing you to identify the domain and context-related opinion terms as well as their orientations. The corpus-based approach aims to provide dictionaries that are specially related to a particular domain (Kharde and Sonawane 2016). To expand the dictionary, statistical or semantic approaches may be used to look for words that are similar as discussed below.

i. *Statistical approach*

The statistical approach includes searching co-occurrence patterns or seed opinion words. Searching for co-occurrence trends or seed opinion terms is one statistical technique. If the corpus is insufficient, the issue of certain words not being available can be solved by using the whole collection of indexed documents on the web as the corpus for creating the dictionary (Turney 2002). In a broad annotated corpus, even the appearance of a word in the positive or negative text may be used to determine its polarity (Read and Carroll 2009). Similar opinion words are likely to co-occur in a corpus, according to Cambria



**Table 2** Lexicons for sentiment analysis

S. no.	Sentiment lexica	Main feature	Pros and/or cons
1	Loughran and McDonald Sentiment Word Lists (Loughran and McDonald 2015)	<ul style="list-style-type: none"><li>• The dictionary reports counts, the proportion of the total, the average proportion per document, the standard deviation of proportion per document, document count, seven sentiment category identifiers, the number of syllables, and the source for each word.</li><li>• Indicator for sentiment related to financial context: “negative”, “positive”, “litigious”, “uncertainty”, “constraining”, or “superfluous”.</li></ul>	Cons: <ul style="list-style-type: none"><li>• Does not contain acronyms, hyphenated words, names, or phrases, British English,</li><li>• Contains a limited number of abbreviations.</li></ul>
2	Stock Market Lexicon (Oliveira et al. 2014)	<ul style="list-style-type: none"><li>• Learning stock market lexicon from StockTwits for the stock market and general financial applications.</li><li>• About 17.44% of the StockTwits messages are labeled as “bullish” or “bearish” by their authors, to show their sentiment toward the mentioned stocks.</li></ul>	Pros: <ul style="list-style-type: none"><li>• Presents Sentiment oriented word embeddings for the stock market.</li></ul> Cons: <ul style="list-style-type: none"><li>• Imbalanced dataset Bullish is much higher than that labeled as Bearish, with an overall ratio of 4.03.</li></ul>
3	SentiWordNet 3.0 (Baccianella et al. 2010)	<ul style="list-style-type: none"><li>• Makes use of WordNet 3.0 to assign positive, negative, and objective scores to terms.</li><li>• Comprises more than 100,000 words that occur in different contexts.</li></ul>	Pros: <ul style="list-style-type: none"><li>• For machine learning based sentiment classification a mixture of documents of different domains achieves good results.</li></ul> Cons: <ul style="list-style-type: none"><li>• For Cross-domain sentiment analysis, rule-based approaches with fixed opinion lexica are unsuited.</li></ul>
4	SenticNet 7 (Cambria et al. 2022)	<ul style="list-style-type: none"><li>• The input sentence is translated from natural language into a sort of “protolanguage” sentence, which generalizes words and multiword expressions in terms of primitives and, hence, connects these (in a semantic-role-labeling fashion) to their corresponding emotion and polarity labels.</li></ul>	Pros: <ul style="list-style-type: none"><li>• Sentence, which generalizes words contains multiword expressions which enable polarity disambiguation.</li></ul> Cons: <ul style="list-style-type: none"><li>• Sentence, which generalizes words do not handle sarcasm or antithetic opinion targets perfectly.</li></ul>

Table 2 (continued)

S. no.	Sentiment lexica	Main feature	Pros and/or cons
5	VADER (Hutto and Gilbert 2014)	<ul style="list-style-type: none"> <li>• A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.</li> <li>• Used for sentiment analysis of text which is sensitive to both the polarities, i.e., positive/negative and finds the intensity (strength) of emotion.</li> <li>• Especially attuned to microblog-like contexts.</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Not only presents the positivity and negativity score but also tells us about how positive or negative a sentiment is.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• May not work for complex data, does not recognize context, and requires additional tools for visualizing output.</li> </ul>
6	Opinion Lexicon (Hu and Liu 2004)	<ul style="list-style-type: none"> <li>• A list of positive and negative opinion words or sentiment words for English customer reviews (around 6800 words).</li> </ul>	<p>Cons:</p> <ul style="list-style-type: none"> <li>• Does not help to find features that are liked by customers.</li> </ul>
7	MPQA Subjectivity Lexicon (Wilson and Wiebe 2003)	<ul style="list-style-type: none"> <li>• In the corpus, individual expressions are marked that correspond to explicit mentions of private states, speech events, and expressive subjective elements.</li> <li>• Annotators were asked to judge all expressions in context.</li> <li>• Includes 5,097 negative and 2,533 positive words. Each word is assigned a strong or weak polarity.</li> <li>• Works great for short sentences, such as tweets or Facebook posts.</li> </ul>	<p>Cons:</p> <ul style="list-style-type: none"> <li>• It is rooted in the subjective interpretations of a single person.</li> </ul>
8	NRC Hashtag Sentiment Lexicon (Mohammad and Kiritchenko 2015; Mohammad 2012)	<ul style="list-style-type: none"> <li>• Association of words with positive (negative) sentiment generated automatically from tweets with sentiment-word hashtags such as #amazing and #terrible.</li> <li>• Number of terms: 54,129 unigrams, 316,531 bigrams, 308,808 pairs,</li> <li>• Association scores: real-valued.</li> </ul>	<p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>

Table 2 (continued)

S. no.	Sentiment lexica	Main feature	Pros and/or cons
9	NRC Hashtag Emotion Lexicon (Mohammad et al. 2013; Zhu et al. 2014; Kiritchenko et al. 2014a)	<ul style="list-style-type: none"> <li>• Association of words with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).</li> <li>• Manually annotated on Amazon's Mechanical Turk.</li> <li>• Generated automatically from tweets with emotion-word hashtags such as #happy and #anger.</li> <li>• Number of terms: 16,862 unigrams (words), 25,000 word senses, Association scores: real-valued</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Available in 40 different languages.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>
10	NRC Hashtag Affirmative Context Sentiment Lexicon (Mohammad et al. 2013; Zhu et al. 2014; Kiritchenko et al. 2014a)	<ul style="list-style-type: none"> <li>• Association of words with positive (negative) sentiment in affirmative or negated contexts generated automatically from tweets with sentiment-word hashtags such as #amazing and #terrible.</li> <li>• Number of terms: Affirmative contexts: 36,357 unigrams, 159,479 bigrams.</li> <li>• Association scores: real-valued</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Available in 40 different languages.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>
11	NRC Hashtag Negated Context Sentiment Lexicon (Mohammad et al. 2013; Zhu et al. 2014; Kiritchenko et al. 2014a)	<ul style="list-style-type: none"> <li>• Association of words with positive (negative) sentiment in negated contexts generated automatically from tweets with sentiment-word hashtags such as #amazing and #terrible.</li> <li>• Number of terms: Negated contexts: 7,592 unigrams, 23,875 bigrams.</li> <li>• Association scores: real-valued</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Available in 40 different languages.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>

Table 2 (continued)

S. no.	Sentiment lexica	Main feature	Pros and/or cons
12	NRC Word-Emotion Association Lexicon/NRC Emotion Lexicon (Mohammad and Turney 2013, 2010)	<ul style="list-style-type: none"> <li>• Association of words with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive) manually annotated on Amazon's Mechanical Turk.</li> <li>• Available in 40 different languages.</li> <li>• Number of terms: 14,182 unigrams (words), ~25,000 word senses.</li> </ul> <p>Association scores: binary (associated or not).</p>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Available in 40 different languages.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>
13	Emoticon Lexicon/Sentiment140 Lexicon (Mohammad et al. 2013; Zhu et al. 2014; Kiritchenko et al. 2014a)	<ul style="list-style-type: none"> <li>• Association of words with positive (negative) sentiment generated automatically from tweets with emoticons such as:) and:(.</li> <li>• Number of terms: 62,468 unigrams, 677, 698 bigrams, 480,010 pairs.</li> <li>• Number of terms: 14,182 unigrams (words), ~25,000 word senses.</li> <li>• Association scores: real-valued.</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Available in 40 different languages.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>
14	Sentiment140 Affirmative Context Lexicon (Mohammad et al. 2013; Zhu et al. 2014; Kiritchenko et al. 2014a)	<ul style="list-style-type: none"> <li>• Association of words with positive (negative) sentiment in affirmative contexts generated automatically from tweets with emoticons such as:) and:(.</li> <li>• Number of terms: Affirmative contexts: 45,255 unigrams, 240,076 bigrams.</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• Available in 40 different languages.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Words can have multiple meanings and senses, and the meaning and sense that is common in one domain may not be common in another. Furthermore, words that are not generally considered sentiment-bearing can imply sentiments in specific contexts.</li> </ul>
15	Yelp Restaurant Sentiment Lexicon (Kiritchenko et al. 2014b)	<ul style="list-style-type: none"> <li>• The Yelp dataset is a subset of our businesses, reviews, and user data for use in personal, educational, and academic purposes.</li> <li>• Created from the Yelp dataset, from the subset of entries about these restaurant-related businesses.</li> </ul>	<p>Cons:</p> <ul style="list-style-type: none"> <li>• Few reviews are considered to be fake.</li> <li>• No proper boundary to detect neutrality.</li> </ul>

Table 2 (continued)

S. no.	Sentiment lexica	Main feature	Pros and/or cons
		<ul style="list-style-type: none"> <li>• Consists of 10 attributes, namely, unique Business ID, Date of Review, Review ID, Stars given by the user, Review given by the user, Type of text entered—Review, Unique User ID, Cool column: The number of cool votes the review received, Useful column: The number of useful votes the review received, Funny Column: The number of funny votes the review received.</li> <li>• Number of reviews: 183,935 reviews.</li> <li>• Have one to five-star ratings associated with each review.</li> </ul>	
16	Amazon Laptop Sentiment Lexicon (McAuley and Leskovec 2013)	<ul style="list-style-type: none"> <li>• Collected reviews posted on Amazon.com from June 1995 to March 2013. Extracted from this subset are all reviews that mention either a laptop or notebook.</li> <li>• Have one to five-star ratings associated with each review.</li> <li>• 26,577 entries for unigrams (includes affirmative and negated context entries), 155,167 entries for bigrams.</li> </ul>	Cons: <ul style="list-style-type: none"> <li>• May not work well with the neutral sentiment.</li> </ul>
17	Macquarie Semantic Orientation Lexicon (Mohammad et al. 2009)	<ul style="list-style-type: none"> <li>• 76,400 terms.</li> <li>• Sentiments: negative, positive</li> <li>• Automatic: Using the structure of a thesaurus and affixes.</li> </ul>	–
18	Harvard's General Inquirer Lexicon (Stone and Hunt 1963)	<ul style="list-style-type: none"> <li>• A lexicon attaching syntactic, semantic, and pragmatic information to part-of-speech tagged words.</li> <li>• 2000 positive and 2000 negative words.</li> </ul>	–

Table 2 (continued)

S. no.	Sentiment lexica	Main feature	Pros and/or cons
19	IMDB (Yenter and Verma 2017)	<ul style="list-style-type: none"> <li>• 50K movie reviews.</li> <li>• A set of 25,000 highly polar movie reviews for training and 25,000 for testing.</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• The data is refreshed daily.</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• IMDB reviews are not considered to be overly trustworthy, as big Hollywood studios generally dictate the scores and the overall consensus.</li> <li>• The algorithm used by IMDB to collate its reviews is generally considered inferior to those used by Rotten Tomatoes and similar sites.</li> </ul>
20	AFINN (Nielsen 2017)	<ul style="list-style-type: none"> <li>• AFINN is the simplest yet most popular lexicon used for sentiment analysis developed by Finn Årup Nielsen.</li> <li>• It contains 3300+ words with a polarity score associated with each word.</li> <li>• A list of English terms manually rated for valence with an integer between -5 (negative) and +5 (positive) by Finn Årup Nielsen between 2009 and 2011.</li> <li>• Primarily analyze Twitter sentiment.</li> </ul>	<p>Cons:</p> <ul style="list-style-type: none"> <li>• Using the raw AFINN score the longer texts may yield higher values simply because they contain more words.</li> </ul>
21	Corpus of Business News (Moreno-Ortiz et al. 2020)	<ul style="list-style-type: none"> <li>• Covers non-specific sentiment-carrying terms and phrases.</li> <li>• It contains 6,470 entries, both single and multi-word expressions, each with tags denoting their semantic orientation and intensity.</li> </ul>	<p>Pros:</p> <ul style="list-style-type: none"> <li>• A wide coverage, a domain-specific lexicon for the analysis of economic and financial texts in English.</li> </ul>
22	DepecheMood Affective Lexicon (Staiano and Guerini 2014)	<ul style="list-style-type: none"> <li>• Harvested crowd-sourced affective annotation from a social news network.</li> <li>• Considered the affective dimensions namely Afraid, Amused, Angry, Annoyed, Don't Care, Happy, Inspired, and Sad.</li> <li>• 37 thousand terms annotated with emotion scores.</li> </ul>	<p>Cons:</p> <ul style="list-style-type: none"> <li>• Cannot handle similar words which are not present in the training document.</li> </ul>

Table 2 (continued)

S. no.	Sentiment lexica	Main feature	Pros and/or cons
23	Financial Phrasebank (Malo et al. 2014)	<ul style="list-style-type: none"><li>• Polar sentiment dataset of sentences from financial news.</li><li>• The dataset consists of 4840 sentences from English-language financial news categorized by sentiment. The dataset is divided by an agreement rate of 5–8 annotators.</li></ul>	<p>Pros:</p> <ul style="list-style-type: none"><li>• Works well for NLP-related tasks in multi-class financial domain classifications.</li></ul>



and Hussain (2015), and hence the polarity of an unfamiliar word can be calculated using the relative frequency of co-occurrence with another word. In this case, PMI can be used (Turney 2002). The statistical approach to the semantic orientation of a word is used in conjunction with PMI (Cambria and Hussain 2015). Another such approach is Latent Semantic Analysis (LSA) (Deerwester et al. 1990).

ii *Semantic approach*

Semantically close words are assigned similar polarities based on this approach. This method is based on various criteria for measuring word similarity (Cambria and Hussain 2015). The relative count of positive and negative synonyms of an unknown word can be used to find out the polarity of that word using different semantic relationships given by WordNet (Kim and Hovy 2004).

A combination of both statistical and semantic approaches is also followed by a few researchers to perform sentiment analysis. Zhang et al. (2012) applied a mixture of both these approaches to online reviews to determine the weakness of products. Sentence-based sentiment analysis, according to their model, is carried out by taking into account the effect of degree adverbs to determine the polarity of each aspect within a sentence. To find the implicit features, they used the collocation statistics-based selection method-Pointwise Mutual Information (PMI). With the use of semantic methods, feature words of the products are grouped into corresponding aspects.

Ding et al. (2008) demonstrated that the same term can have multiple polarities in different contexts, even within the same domain. Therefore, rather than simply finding domain-dependent sentient words using the corpus-based approach, they explored the notion of intra-sentential and inter-sentential sentiment consistency.

In the lexicon-based approach, one point is worth noticing. The initial manual annotation of the seed list can be a costly procedure. Secondly and most importantly, the use of a dictionary even for seed list generation can lead to the insufficiency of handling cross-domain problems. Thus, the usage of a proper technique to generate a seed list for a lexicon-based approach is an open problem. Also, whenever linguistic rules are involved in handling knowledge, there might be situations where it fails to correctly grasp the affective sentiment.

Hybrid approaches which use sentiment lexicons in machine learning methods have also attracted many researchers to combine the benefits of both approaches. Trinh et al. (2018) used the hybrid approach to perform sentiment analysis of Facebook comments in the Vietnamese language. While their dictionary is partly based on SO-CAL, the authors manually built the dictionary to include nouns, verbs, adjectives, and adverbs along with emotional icons. They performed sentence-level sentiment analysis of product reviews using the SVM classifier. Appel et al. (2016) also performed sentence-level sentiment analysis using a combination of lexicon and machine learning approaches. They extended their sentiment lexicon with the use of SWN and used fuzzy sets to determine the polarity of sentences. Using an SVM classifier, Zhang et al. (2011) performed entity-level sentiment analysis of tweets, with the use of a lexicon that supports business marketing or social studies. They made use of the lexicon by Ding et al. (2008) along with some frequently used opinion hashtags to build the lexicon for their model. Pitogo and Ramos (2020) performed sentiment analysis for Facebook comments using a lexicon-based approach called

Valence Aware Dictionary and Sentiment Reasoner (VADER) along with a hierarchical clustering algorithm.

### 3.5 Sentiment or opinion summarization

Sentiment or Opinion summarization or aggregation aims to provide an idea of the overall influence or polarity depicted by the dataset, by summing up the polarity of all individual words/aspects /sentences/documents of the dataset. Sentiment summarization must not be confused with text summarization, though they are slightly related. Text summarization aims to provide a summary of the dataset, while sentiment summarization provides a generalized polarity depicted by the whole dataset.

Different types of summarization models are proposed by researchers to obtain an average sentiment. Pang and Lee (2004) first extracted all subjective sentences and then summarized those subjective sentences. Blair-Goldensohn et al. (2008) used a tool to choose a few representative documents from a vast number of documents and then used them for emotion summarization based on aspects. By mining opinion features from product feedback, Hu and Liu (2004) suggested an aspect-based sentiment summarization strategy for online consumer reviews. Using the ratings on different aspects, Titov and McDonald (2008) proposed a model which can contribute to the sentiment summarization process. Their algorithm is designed to find related topics in text and collect textual evidence from reviews to support aspect ratings. Sokolova and Lapalme (2009) developed an emotion summarization model to summarise the opinionated text in consumer goods by integrating different polarity detection techniques and automated aspect detection algorithms. Different types of summarization models are proposed by researchers to obtain an average sentiment. Pang and Lee (2004) first extracted all subjective sentences and then summarized those subjective sentences. Blair-Goldensohn et al. (2008) used a tool to choose a few representative documents from a vast number of documents and then used them for emotion summarization based on aspects. By mining opinion features from product feedback, Hu and Liu (2004) suggested an aspect-based sentiment summarization strategy for online consumer reviews. Using the ratings on different aspects, Titov and McDonald (2008) proposed a model which can contribute to the sentiment summarization process. Their algorithm is designed to find related topics in text and collect textual evidence from reviews to support aspect ratings. Bahrainian and Dengel (2013) developed an emotion summarization model to summarise the opinionated text in consumer goods by integrating different polarity detection techniques and automated aspect detection algorithms.

## 4 Performance analysis measures

The evaluation of performance is one of the principal concepts associated with building a resourceful model. Once the sentiments are classified as either positive or negative, the performance of the model needs to be evaluated. The papers by Sokolova and Lapalme (2009) provided a better understanding of the applicability of performance measures depending on the variability of the classification tasks. Among different kinds of available metrics for measuring the performance of a textual sentiment analysis model, metrics based on the confusion matrix are widely used (Sokolova and Lapalme 2007, 2009; John and Kartheeban 2019). The details concerning the classifications that are expected and those that are calculated by a classifier are

**Table 3** Confusion matrix for binary classification

		Actual classification	
		Positive	Negative
Expectation	Positive	TP	FN
	Negative	FP	TN

*TP* overall positive data that are classified as positive, *TN* overall negative data that are classified as negative, *FP* overall negative data that are classified as positive, *FN* overall positive data that are classified as negative

shown in the confusion matrix. A confusion matrix for binary classification problems consists of four separate data entries, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), as shown in Table 3.

The most frequently used performance metric is accuracy to measure the overall effectiveness of the model. Accuracy determines the proportion of a total number of instances (i.e., documents/ sentences/words) that are correctly predicted by the sentiment analysis model. Equation 13 shows the formula for estimating the model's accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

Apart from accuracy, precision and recall are well-known metrics that are best suited for text applications (Sokolova and Lapalme 2007). The number of correctly classified positive instances is determined by positive predictive value or precision, while the number of correctly classified negative instances is determined by negative predictive value. The proportion of actual positive instances that are correctly classified is determined by sensitivity or recall; the proportion of actual negative instances that are correctly classified is determined by negative recall or specificity.

The following are the formulas for calculating them (Salari et al. 2014).

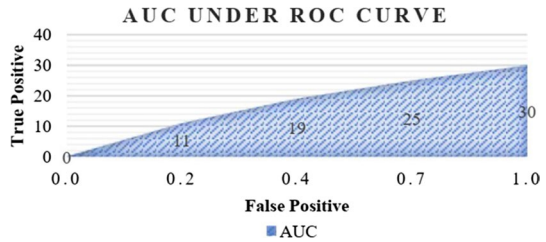
$$Precision/Positive predictive value = \frac{TP}{TP + FP} \quad (14)$$

$$Negative predictive value = \frac{TN}{TN + FN} \quad (15)$$

$$Recall/Sensitivity = \frac{TP}{TP + FN} \quad (16)$$

$$Specificity/Negative recall = \frac{TN}{TN + FP} \quad (17)$$

Precision and recall are better indicators of the current system's success than accuracy for an imbalanced binary classifier. Yet, in certain situations, a system may have high precision but poor recall, or vice versa. In this case, the f-measure allows you to articulate all issues with a single number. Once the precision and recall for a binary or multi-class classification task have been calculated, the two scores together form the f-measure, as seen in Eq. 18.

**Fig. 3** AUC under ROC

$$F\text{-measure}, F = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

Accuracy or f-measure can show overoptimistic inflated results, especially on imbalanced datasets. Matthew's Correlation Coefficient (MCC) is a more reliable statistical rate that produces a high score only if the prediction obtained good results in all of the four confusion matrix categories proportionally, both to the size of positive elements and the size of negative elements in the dataset. The confusion matrix or an error matrix can be summed up using MCC as shown in Eq. 19. MCC ranges from  $[-1, 1]$ , where 1 indicates the best agreement between the predicted and actual values. The MCC helps us to identify the ineffectiveness of the classifier in classifying especially the minority class samples.

$$MCC = \frac{TN * TP - FN * FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (19)$$

To measure the ability of a sentiment classifier to distinguish between the polarity classes, an Area Under the Curve (AUC) is employed. The curve in AUC is generally a ROC (Receiver Operating Characteristic) curve, which is a graph showing the performance of a classification model at all classification thresholds as shown in Fig. 3. ROC plots TP and FP. AUC is an aggregated evaluation of the classifier as the threshold varies over all possible values. The Precision-Recall AUC summarizes the curve using a range of threshold values as a single score. AUC measures how true positive rate (recall) and false positive rate trade-off. Specifically, for imbalanced datasets, where overfitting needs to be avoided, AUC works as a preferable evaluation matrix. AUC represents the probability that a random positive instance is positioned to the right of a random negative instance. AUC ranges from 0 to 1. An AUC of 0.0 denotes a model that makes all incorrect classifications, whereas an AUC of 1.0 denotes a model that makes all correct classifications.

When a regression task is adopted for sentiment analysis, Mean Squared Error (MSE) is employed to find the squared difference between actual and predicted values. It is an absolute measure of the goodness of fit of dependent variables in the model. The formula of MSE is given in Eq. 20. The lower the value of MSE, the better the sentiment analyzer. It can be used as a loss function as the graph of MSE is differentiable. However, it is not very suitable in case the dataset contains outliers.

$$MSE = \frac{1}{n} \left( \sum (y - \hat{y})^2 \right) \quad (20)$$

In contrast to the context dependency of MSE, R squared is a context-independent metric that is used for a regression task. It is a relative measure of how well the model fits dependent variables or how close the data is to the fitted regression line. Coefficient of

Determination and Goodness of Fit are other names for  $R$  squared and it is calculated using Eq. 21.

$$R^2 = 1 - \frac{SSR}{SSM} \quad (21)$$

where  $SSR$  is the squared sum error of the regression line and  $SSM$  is the squared sum error of the mean line.

Other performance evaluation metrics that can be also considered for evaluating a sentiment analysis model are Root Mean Squared Error (RMSE), Residual Standard Error (RSE), Mean Absolute Error (MAE), etc.

## 5 Applications of sentiment analysis

Sentiment analysis or opinion mining has recently been used in studies on e-commerce feedback, tweets, Facebook posts, YouTube content, blog entries, and a variety of other data mining and knowledge-based AI programs. As a result, it has progressed significantly in fields including Information Retrieval (IR), web data analysis, mining of text, analysis of text, NLP, computational linguistics, and biometrics. Using different approaches/methods/frameworks analyzed in this paper beforehand, sentiment analysis can be applied to various fields such as tourism, education, defense, business, politics, public, finance, hazards, health, and safety. The broad range of applications will aim to obtain the best possible combination of strengths, whether or not any of the components in Fig. 1 or any of the approaches indicated in Fig. 2 are present. Depending on the requirement/aim/ framework of a sentiment analysis model, applications can vary from a straightforward prediction of the polarity of a single word to uncovering sensitive or hidden information, or even a pattern to protect a nation from any potential terrorist attack or disaster. Many research works mention different application areas based on different domains or approaches used (Alessia et al. 2015; Jain and Gupta 2022; Saxena et al. 2022; Feldman 2013; Govindarajan 2022; Ravi and Ravi 2015). The knowledge of diverse application fields based purely on the dataset at hand is challenging to find in existing research papers. This paper aims to outline several sentiment analysis application areas based on the data/content/material in hand, that can be used by researchers for sentiment analysis.

### 5.1 Reviews on products

Sentiment analysis using reviews on different products with different brands is the most widespread practice, which encompasses different application angles. For a particular product, the number of brands has been increasing day to day. Also, the same brand may offer products with different specifications. Nowadays even different online shopping sites are available that sell the same product. This creates confusion among customers to reach an optimal decision. Though shopping sites offer the option of displaying comments and star ratings left by former customers to assist potential buyers, the count of current feedback can be so large that scrolling through thousands of them can be a time-consuming process. Sentiment analysis helps to alleviate this condition by giving a concise perspective on a product or brand as a whole, or even on a certain feature/aspect of the product. Also, it can be used by the sellers or manufacturers to concentrate on the suitable aspects or specifications, which can be used for upgrading the product or deciding the advertisement

strategy. Product analysis by buyers, suppliers, and sellers; competitor analysis or market study by sellers or manufacturers; brand tracking and reputation management by manufacturers; customer service by e-commerce sites; and customer analysis by sellers and manufacturers are among the various application directions associated with sentiment analysis of product feedback. The necessity to detect fake reviews before using the available data for decision-making was highlighted in the research work by Vidanagama et al. (2022). The authors made use of a rule-based classifier, a domain feature ontology, and Mahalanobis distance to detect fake reviews while performing aspect-based sentiment analysis. Cao et al. (2022) have introduced a quality evaluation model of products by combining deep learning, word vector conversion, keyword clustering, and feature word extraction technologies. Their model improves product features based on consumer online reviews and finally calculates customer satisfaction and attention based on short text comments with sentiment tags. With the use of pre-trained word embeddings, Bhuvaneshwari et al. (2022) proposed a Bi-LSTM Self Attention based CNN (BAC) model for analysis of user reviews. Wang et al. (2022b) designed multi-attention bi-directional LSTM (BLSTM(MA)), and used Latent Dirichlet Allocation (LDA) modeling to perform multimodal fusion for sentiment analysis of product reviews. Alantari et al. (2022) examined 260,489 reviews from five review platforms, covering 25,241 products in nine different product categories. They discovered that pretrained neural network-based machine learning techniques, in particular, provide the most precise forecasts, while topic models like LDA provide more thorough diagnostics. To make predictions, topic models are better suited than neural network models, which are not good at making diagnoses. As a result, the preference of the analysts for prediction or diagnostics is likely to determine how text review processing technologies are chosen in the future.

## 5.2 Political Tweets, Facebook comments, Blog posts, and YouTube Videos

Recently, people have started to openly share their views or opinion on different political parties, electoral candidates, government policies, and rules on different public platforms such as Twitter, Facebook, YouTube, and blogs. These create a great influence on the followers. Therefore, they are used by many experts to predict the outcome of an election beforehand, monitor public sentiment on various political movements, or analyze the sentiment of the public on a proposed government rule, bill, or law.

With the use of pre-trained models and the Chi-square test, Antypas et al. (2022) proposed a multilingual sentiment analysis model to analyze both influential and less popular politicians' tweets from members of parliament of Greece, Spain, and the United Kingdom. Their study indicates that negative tweets spread rapidly as compared to positive ones. Using Valence Aware Dictionary and sentiment Reasoner (VADER), and 2 million tweets on the 2019 Indian Lok Sabha Election, Passi and Motisariya (2022) analyzed sentiments of Twitter users towards each of the Indian political parties. Using the aging estimation method with the proportion of positive message rate to negative messages rate, Yavari et al. (2022) designed an indicator of the election results in the future.

## 5.3 Tweets or comments on Facebook/YouTube/Instagram on social cause or events

Expressions of opinions on different social causes or events have also increased recently. This increases the scope of designing application portals that perform public sentiment analysis, monitor, and predict different possible outcomes of such an event or cause and

decide the possible steps which need to be adopted in the future in case there is an outbreak of any chaotic situation.

A multi-grained sentiment analysis and event summary method employing crowd-sourced social media data on explosive accidents was built by Ouyang et al. (2017). The system can determine which components of the event draw users' attention, identify which microblog is responsible for a large shift in sentiment, and detect those aspects of the event that affect users' attention. Smith and Cipolli (2022) studied the emotional discourse before and after a prohibition on graphic photos of self-harm on Facebook and Instagram using a corpus of 8,013 tweets. By clarifying topical content using statistical modeling to extract abstract topics in discourse, the authors offered an insight into how the policy change relating to self-harm was viewed by those with a vested interest.

#### **5.4 Reviews on Blogs/Tweets/Facebook comments on movie**

Reviews on an upcoming movie or a movie that is onscreen in the theatres can be used to decide the success or failure of the movie. Different movie recommender systems can also be designed using the reviews from the audience. Also, the distributors or producers can use such reviews to improve their advertising strategy based on the different aspects which are liked by the viewers.

Using sentiment analysis to gain a deeper understanding of user preferences, Dang et al. (2021) proposed methods to enhance the functionality of recommender systems for streaming services. The Multimodal Album Reviews Dataset (MARD) and Amazon Movie Reviews were used to test and compare two different LSTM and CNN combinations, LSTM-CNN and CNN-LSTM. They started with a version of the recommendation engine without sentiment analysis or genres as their baseline. As compared to the baseline, the results demonstrate that their models are superior in terms of rating prediction and top recommendation list evaluation. Pavitha et al. (2022) designed a system for analyzing movie reviews in different languages, classifying them into either positive or negative using Naive Bayes and Support Vector Classifier (SVC), and recommending similar movies to users based on Cosine Similarity. For B-T4SA and IMDB movie reviews, Zhu et al. (2022) proposed a self-supervised sentiment analysis model namely Senti-ITEM. The model pairs a representative image with the social media text as a pretext task, extract features in a shared embedding space, and uses SVM for sentiment classification.

#### **5.5 Tweets/Facebook comments on pandemic/crisis /environmental issues**

Nowadays people encountering abrupt situations or difficulties due to the Covid-19 pandemic or any environmental issues such as storm or earthquake posts real-time tweets or comments on Facebook. In such a situation, by analyzing tweets or comments properly, government or any agency, or even nearby people can offer help, and perform disaster management and crisis analysis.

Hodson et al. (2022) suggested a corpus-assisted discourse analysis approach, for analyzing public opinion on COVID-19 tweets and YouTube comments related to Canadian Public Health Office. The authors found that different platforms convey key differences between comments, specifically based on the tone used in YouTube videos as compared to plain text in Tweets. To capture sarcasm or get clear information, cross-platform and diverse methods must be adopted to facilitate health-related communication and public opinion. Chopra et al. (2022) employed logistic regression, Naive Bayes, XGBoost, LSTM,



GloVe, and BERT to predict disaster warnings from tweets and evaluate the seriousness of the content.

## 5.6 Tweets/Facebook comments/YouTube videos on Stock Market

One of the trending application areas of sentiment analysis is Stock Market Prediction. Identifying stocks and share with great potential and deciding the optimal time to buy them at the lowest price and sell them at the peak time can be performed using a suitable sentiment analysis model. Using stock market data with SVM, Ren et al. (2018) suggested a model that forecasts movement direction and predicts stock prices while capturing investor psychology. Sousa et al. (2019) used the BERT algorithm to analyze the sentiments of news articles and provide relevant information that can facilitate stock market-related quick decision-making. Considering both positive and negative financial news, de Oliveira Carosia et al. (2021) analyzed the influence on the stock market using three Artificial Deep Neural Networks namely Multi-Layer Perceptron (MLP), LSTM, and CNN. The findings of this sentiment analysis model's observations revealed that while recurrent neural networks, such as LSTM, perform better in terms of time characteristics when used to predict the stock market, CNNs perform better when assessing text semantics.

## 6 Future scope of research in sentiment analysis

There are numerous scientific studies in the literature that focus on each of the components of the sentiment analysis approach, either independently or in combination. Each of these sentiment analysis modules offers plenty of opportunities for further investigation, improvisation, and innovation. Several challenges and issues are also faced during the process of performing sentiment analysis, which may hinder the proper functioning or performance of the model. Some of them are domain dependency, reference problems, sarcasm detection, spam detection, time period, etc. Most of these challenges influence the development of better techniques and algorithms to handle them. Some of the primary research gaps that offer scope for future research and hence encourage further sentiment analysis research are discussed below:

1. It has been found that current techniques dedicated to sentiment analysis do not employ effective data initialization and pre-processing techniques. Rather than relying on established NLP pre-processing techniques, an advanced pre-processing technique, such as standard normalization that takes deliberately into account, the case of negation and mixed emotion would be extremely beneficial.
2. One of the most critical steps in improving the performance of a sentiment analysis model is keyword extraction. Many sentiment analysis models have been observed to extract keywords using generalized dictionaries. The use of generalized dictionaries, on the other hand, produces inaccurate findings since most of these dictionaries include keywords that are relevant to a specific domain. However, there is no predefined list of keywords for a certain domain or subject in the real world. Different researchers have shown the supremacy of the degree centrality metric for the graph-based method of obtaining the best collection of representative and sentimental words. As a result, it may be used to find key terms or phrases. Automatic keyword extraction techniques can be used for sentiment analysis in a variety of applications, both independently and in

- combination. Most of these techniques have found applications in a variety of research areas, including Data Analysis, TM, IR, and NLP since they allow for the condensing of text records.
3. Assignment of polarity scores to keywords using sentiment dictionaries has gained a lot of attention in sentiment analysis. However, depending on its use in a specific domain, a term can serve as a positive or negative word at different times. Therefore, the usage of sentiment dictionaries with pre-defined polarities for words is not an appropriate practice for sentiment analysis. Existing sentiment dictionaries fail to handle sarcasm or negations to a great extent. It is observed that many machine learning based techniques are trained to work only for a particular domain. They do not consider that the words can change their polarity based on the context and domain of application. Thus, whenever the same word is tested for another domain using the trained classifier, it shows incorrect results in some situations.
  4. New edge and node weighing approaches may be introduced and used in place of NE-Rank or TextRank centralities to determine keyword rank. To achieve improved outcomes in the future, different ensemble or individual improvised centralities may be used. This establishes a framework for future research into graph mining algorithms for sentiment analysis in various fields.

## 7 Conclusion

The era of digitization marks the astonishing growth of subjective textual data online. Proper analysis of the textual information, to rightly reflect the public sentiment regarding any topic, demands proper investigation of textual data. Sentiment analysis has emerged as the most important task which helps to enhance the decision-making process by extracting the underlying sentiment or opinion of data. Even though sentiment analysis has progressed in recent years, modern models have flaws such as domain dependence, negation management, high dimensionality, and the failure to use efficient keyword extraction. This paper examines and provides a comprehensive discussion of different perspectives related to the creation and implementation of an effective sentiment analysis model. A thorough examination and establishment of various modules of the sentiment analysis methodology are carried out to plan and improve effective sentiment analysis models. The keyword extraction algorithm is vital to the success of a sentiment analysis model and thus is well-studied in this paper. The paper also discusses sentiment classification methods, which form an essential aspect of a sentiment analysis model. The paper conducts a detailed review of both machine learning and lexicon-based approaches to textual data sentiment analysis.

As a thorough, well-organized study on sentiment analysis, this research effort can assist academicians and industry experts in analyzing and developing powerful sentiment analysis models in a wide range of domains. Sentiment analysis models have a lot of potential for further development and use in the near future because they have a broad range of uses in social, industrial, political, economic, health and safety, education, defense financial contexts, and others. Each of the sentiment analysis modules as discussed in this paper can be investigated, improvised, and supplemented with certain relevant algorithms to design an efficient sentiment analysis model. This study also offers prospective guidelines for carrying out proper sentiment analysis research.

## References

- Abilhoa WD, De Castro LN (2014) A keyword extraction method from twitter messages represented as graphs. *Appl Math Comput* 240:308–325
- Ahmad M, Aftab S, Bashir MS, Hameed N (2018) Sentiment analysis using SVM: a systematic literature review. *Int J Adv Comput Sci Appl* 9(2)
- Alantari HJ, Currin IS, Deng Y, Singh S (2022) An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. *Int J Res Mark* 39(1):1–19
- Alessia D, Ferri F, Grifoni P, Guzzo T (2015) Approaches, tools and applications for sentiment analysis implementation. *Int J Comput Appl* 125(3):26–33
- Alfter D, Cardon R, François T (2022) A dictionary-based study of word sense difficulty. In: Proceedings of the 2nd workshop on tools and resources to empower people with READING Difficulties (READI) within the 13th language resources and evaluation conference. European Language Resources Association, pp 17–24
- Altheneyan AS, Menai MEB (2014) Naïve bayes classifiers for authorship attribution of Arabic texts. *J King Saud Univ-Comput Inf Sci* 26(4):473–484
- Antypas D, Preece A, Collados JC (2022) Politics and virality in the time of twitter: a large-scale cross-party sentiment analysis in Greece, Spain and united kingdom. *arXiv preprint arXiv:2202.00396*
- Appel O, Chiclana F, Carter J, Fujita H (2016) A hybrid approach to the sentiment analysis problem at the sentence level. *Knowl-Based Syst* 108:110–124
- Athanasious V, Maragoudakis M (2017) A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: a case study for modern Greek. *Algorithms* 10(1):34
- Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources and evaluation (LREC'10), vol 10. European Language Resources Association (ELRA), pp 2200–2204
- Bahrainian S-A, Dengel A (2013) Sentiment analysis and summarization of twitter data. In: 2013 IEEE 16th international conference on computational science and engineering. IEEE, pp 227–234
- Baid P, Gupta A, Chaplot N (2017) Sentiment analysis of movie reviews using machine learning techniques. *Int J Comput Appl* 179(7):45–49
- Balahur A, Hermida JM, Montoyo A (2011) Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *IEEE Trans Affect Comput* 3(1):88–101
- Banea C, Mihalcea R, Wiebe J (2014) Sense-level subjectivity in a multilingual setting. *Comput Speech Lang* 28(1):7–19
- Bao H, Li Q, Liao SS, Song S, Gao H (2013) A new temporal and social PMF-based method to predict users' interests in micro-blogging. *Decis Support Syst* 55(3):698–709
- Beliga S (2014) Keyword extraction: a review of methods and approaches. University of Rijeka, Department of Informatics, Rijeka 1(9)
- Beliga S, Meštrović A, Martinčić-Ipšić S (2015) An overview of graph-based keyword extraction methods and approaches. *J Inf Org Sci* 39(1):1–20
- Bellaachia A, Al-Dhelaan M (2012) Ne-rank: a novel graph-based keyphrase extraction in twitter. In: 2012 IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology, vol 1. IEEE, pp 372–379
- Benghuzzi H, Elsheh MM (2020) An investigation of keywords extraction from textual documents using word2vec and decision tree. *Int J Comput Sci Inf Secur* 18(5)
- Bhargav PS, Reddy GN, Chand RR, Pujitha K, Mathur A (2019) Sentiment analysis for hotel rating using machine learning algorithms. *Int J Innov Technol Explor Eng (IJITEE)* 8(6):1225–1228
- Bharti SK, Babu KS (2017) Automatic keyword extraction for text summarization: a survey. *arXiv preprint arXiv:1704.03242* 4:410–427
- Bhuvaneshwari P, Rao AN, Robinson YH, Thippeswamy M (2022) Sentiment analysis for user reviews using bi-lstm self-attention based CNN model. *Multimedia Tools Appl* 81(9):12405–12419
- Blair-Goldensohn S, Hannan K, McDonald R, Neylon T, Reis G, Reynar J (2008) Building a sentiment summarizer for local service reviews. WWW2008 workshop on NLP challenges in the information explosion era
- Blitzer J, Dredze M, Pereira F (2007) Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In: Proceedings of the 45th annual meeting of the Association of Computational Linguistics. ACL, pp 440–447
- Boiy E, Moens M-F (2009) A machine learning approach to sentiment analysis in multilingual web texts. *Inf Retrieval* 12(5):526–558

- Bollegala D, Weir D, Carroll J (2012) Cross-domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Trans Knowl Data Eng* 25(8):1719–1731
- Bonacich P (2007) Some unique properties of eigenvector centrality. *Soc Netw* 29(4):555–564
- Bougouin A, Boudin F, Daille B (2013) Topicrank: graph-based topic ranking for keyphrase extraction. In: *International joint conference on natural language processing (IJCNLP)*, pp 543–551
- Bronselaer A, Pasi G (2013) An approach to graph-based analysis of textual documents. In: *8th European Society for fuzzy logic and technology (EUSFLAT-2013)*. Atlantis Press, pp 634–641
- Cambria E (2013) An introduction to concept-level sentiment analysis. In: Castro F, Gelbukh A, González M (eds) *Advances in soft computing and its applications*. Mexican international conference on artificial intelligence, MICAI 2013. *Lecture notes in computer science*, vol 8266. Springer, pp 478–483
- Cambria E (2016) Affective computing and sentiment analysis. *IEEE Intell Syst* 31(02):102–107
- Cambria E, Hussain A (2015) Sentic computing. *Cogn Comput* 7(2):183–185
- Cambria E, Liu Q, Decherchi S, Xing F, Kwok K (2022) Senticnet 7: a commonsense-based neurosymbolic AI framework for explainable sentiment analysis. In: *Proceedings of LREC 2022*. European Language Resources Association (ELRA), pp 3829–3839
- Cao J, Li J, Yin M, Wang Y (2022) Online reviews sentiment analysis and product feature improvement with deep learning. *Trans Asian Low-Resour Lang Inf Process*. <https://doi.org/10.1145/3522575>
- Castillo E, Cervantes O, Vilarino D, Báez D, Sánchez A (2015) Udlap: sentiment analysis using a graph-based representation. In: *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*. ACL, pp 556–560
- Cepeda K, Jaiswal R (2022) Sentiment analysis on covid-19 vaccinations in Ireland using support vector machine. In: *2022 33rd Irish signals and systems conference (ISSC)*. IEEE, pp 1–6
- Chan JY-L, Bea KT, Leow SMH, Phoong SW, Cheng WK (2022) State of the art: a review of sentiment analysis based on sequential transfer learning. *Artif Intell Rev* 1–32
- Chen P-I, Lin S-J (2010) Automatic keyword prediction using google similarity distance. *Expert Syst Appl* 37(3):1928–1938
- Chen K, Zhang Z, Long J, Zhang H (2016) Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert Syst Appl* 66:245–260
- Chen Y, Wang J, Li P, Guo P (2019) Single document keyword extraction via quantifying higher-order structural features of word co-occurrence graph. *Comput Speech Lang* 57:98–107
- Chen Z, Xue Y, Xiao L, Chen J, Zhang H (2021) Aspect-based sentiment analysis using graph convolutional networks and co-attention mechanism. In: *International conference on neural information processing*. Springer, pp 441–448
- Chong WY, Selvaretnam B, Soon L-K (2014) Natural language processing for sentiment analysis: an exploratory analysis on tweets. In: *2014 4th international conference on artificial intelligence with applications in engineering and technology*. IEEE, pp 212–217
- Chopra M, Singh SK, Aggarwal K, Gupta A (2022) Predicting catastrophic events using machine learning models for natural language processing. In: *Data mining approaches for big data and sentiment analysis in social media*. IGI Global, pp 223–243
- Cortis K, Freitas A, Daudert T, Huerlimann M, Zarrouk M, Handschuh S, Davis B (2017) Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. In: *11th International workshop on semantic evaluations (SemEval-2017): Proceedings of the workshop*, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL), pp 519–535
- Cruz Mata F, Troyano Jiménez JA, de Salamanca Enríquez, Ros F, Ortega Rodríguez FJ, García Vallejo CA (2013) ‘Long autonomy or long delay?’ The importance of domain in opinion mining. *Expert Syst Appl* 40(8):3174–3184
- Dai A, Hu X, Nie J, Chen J (2022) Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis. *Int J Data Sci Anal* 14(1):17–26
- Dang Y, Zhang Y, Chen H (2009) A lexicon-enhanced method for sentiment classification: an experiment on online product reviews. *IEEE Intell Syst* 25(4):46–53
- Dang CN, Moreno-García MN, De la Prieta F (2021) Using hybrid deep learning models of sentiment analysis and item genres in recommender systems for streaming services. *Electronics* 10(20):2459
- Darena F, Zizka J, Burda K (2012) Grouping of customer opinions written in natural language using unsupervised machine learning. In: *2012 14th international symposium on symbolic and numeric algorithms for scientific computing*. IEEE, pp 265–270
- Dave K, Lawrence S, Pennock DM (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: *Proceedings of the 12th international conference on World Wide Web*. ACM, pp 519–528

- de Oliveira Carosia AE, Coelho GP, da Silva AEA (2021) Investment strategies applied to the Brazilian stock market: a methodology based on sentiment analysis with deep learning. *Expert Syst Appl* 184:115470
- Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. *J Am Soc Inf Sci* 41(6):391–407
- Devika R, Subramaniaswamy V (2021) A semantic graph-based keyword extraction model using ranking method on big social data. *Wirel Netw* 27(8):5447–5459
- Devika M, Sunitha C, Ganesh A (2016) Sentiment analysis: a comparative study on different approaches. *Procedia Comput Sci* 87:44–49
- Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT. ACL*, pp 4171–4186
- Ding X, Liu B, Yu PS (2008) A holistic lexicon-based approach to opinion mining. In: *Proceedings of the 2008 international conference on web search and data mining*, vol 39. Elsevier, pp 231–240
- Duari S, Bhatnagar V (2019) sclave: semantic connectivity aware keyword extraction. *Inf Sci* 477:100–117
- Fahrni A, Klenner M (2008) Old wine or warm beer: target-specific sentiment analysis of adjectives. *University of Zurich*, pp 60–63
- Fazlourrahman B, Aparna B, Shashirekha H (2022) Coffitt-covid-19 fake news detection using fine-tuned transfer learning approaches. In: *Congress on intelligent systems*, vol 111. Springer, pp 879–890
- Feldman R (2013) Techniques and applications for sentiment analysis. *Commun ACM* 56(4):82–89
- Gamon M (2004) Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In: *COLING 2004: proceedings of the 20th international conference on computational linguistics. ACL and AFNLP*, pp 841–847
- Ghosal D, Majumder N, Poria S, Chhaya N, Gelbukh A (2019) Dialoguecn: a graph convolutional neural network for emotion recognition in conversation. In: *EMNLP-IJCNLP 2019-2019 conference on empirical methods in natural language processing and 9th international joint conference on natural language processing, proceedings of the conference*, pp 154–164
- Golbeck J (2013) *Analyzing the Social Web*. Newnes, ACM
- Govindarajan M (2022) Approaches and applications for sentiment analysis: a literature review. In: *Data mining approaches for big data and sentiment analysis in social media*. IGI Global, pp 1–23
- HaCohen-Kerner Y (2003) Automatic extraction of keywords from abstracts. In: *International conference on knowledge-based and intelligent information and engineering systems*. Springer, pp 843–849
- Haddi E, Liu X, Shi Y (2013) The role of text pre-processing in sentiment analysis. *Procedia Comput Sci* 17:26–32
- Hart L (2013) *The linguistics of sentiment analysis*. University Honors Theses. <https://doi.org/10.15760/honors.19>
- Hazarika D, Poria S, Zadeh A, Cambria E, Morency L-P, Zimmermann R (2018) Conversational memory network for emotion recognition in dyadic dialogue videos. In: *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, vol 2018. NIH Public Access, p 2122
- Hemalatha I, Varma GS, Govardhan A (2013) Sentiment analysis tool using machine learning algorithms. *Int J Emerg Trends Technol Comput Sci* 2(2):105–109
- Hidayat THJ, Ruldeviyani Y, Aditama AR, Madya GR, Nugraha AW, Adisaputra MW (2022) Sentiment analysis of twitter data related to Rinca island development using doc2vec and SVM and logistic regression as classifier. *Procedia Comput Sci* 197:660–667
- Hidayatullah AF, Cahyaningtyas S, Hakim AM (2021) Sentiment analysis on twitter using neural network: Indonesian presidential election 2019 dataset. In: *IOP conference series: materials science and engineering*, vol 1077. IOP Publishing, p 012001
- Hitesh M, Vaibhav V, Kalki YA, Kamtam SH, Kumari S (2019) Real-time sentiment analysis of 2019 election tweets using word2vec and random forest model. In: *2019 2nd international conference on intelligent communication and computational techniques (ICCT)*. IEEE, pp 146–151
- Hodson J, Veletsianos G, Houlden S (2022) Public responses to covid-19 information from the public health office on twitter and Youtube: implications for research practice. *J Inf Technol Polit* 19(2):156–164
- Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. In: *Proceedings of the 56th annual meeting of the Association for Computational Linguistics*, vol 1. ACL
- Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2):415–425
- Hu Y, Li W (2011) Document sentiment classification by exploring description model of topical terms. *Comput Speech Lang* 25(2):386–403
- Hu M, Liu B (2004) Mining opinion features in customer reviews. In: *AAAI*, vol 4. AAAI, pp 755–760

- Huang C, Zhao Q (2022) Sensitive information detection method based on attention mechanism-based Elmo. *J Comput Appl* 42(7):2009–2014
- Hulth A (2003) Improved automatic keyword extraction given more linguistic knowledge. In: Proceedings of the 2003 conference on empirical methods in natural language processing. ACL, pp 216–223
- Hutto C, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media, vol 8. AAAI, pp 216–225
- Jain AP, Dandannavar P (2016) Application of machine learning techniques to sentiment analysis. In: 2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT). IEEE, pp 628–632
- Jain S, Gupta V (2022) Sentiment analysis: a recent survey with applications and a proposed ensemble algorithm. In: Computational intelligence in data mining. Springer, pp 13–25
- Jain PK, Quamer W, Saravanan V, Pamula R (2022) Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis. *J Ambient Intell Human Comput* 1–13
- Ji X, Chun S, Wei Z, Geller J (2015) Twitter sentiment classification for measuring public health concerns. *Soc Netw Anal Min* 5(1):1–25
- Jianqiang Z (2015) Pre-processing boosting twitter sentiment analysis? In: 2015 IEEE international conference on Smart City/SocialCom/SustainCom (SmartCity). IEEE, pp 748–753
- Jianqiang Z, Xiaolin G (2017) Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access* 5:2870–2879
- Jiao W, Lyu M, King I (2020) Real-time emotion recognition via attention gated hierarchical memory network. Proceedings of the AAAI conference on artificial intelligence 34:8002–8009
- John SM, Kartheeban K (2019) Sentiment scoring and performance metrics examination of various supervised classifiers. *Int J Innov Technol Explor Eng* 9(2S2), 1120–1126
- Joshi M, Rosé C (2009) Generalizing dependency features for opinion mining. In: Proceedings of the ACL-IJCNLP 2009 conference short papers. ACL, pp 313–316
- Kanayama H, Nasukawa T (2006) Fully automatic lexicon expansion for domain-oriented sentiment analysis. In: Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 355–363
- Katrekhar A, AVP BDA (2005) An introduction to sentiment analysis. GlobalLogic Inc 1–6
- Kaur R, Kautish S (2022) Multimodal sentiment analysis: a survey and comparison. In: Research anthology on implementing sentiment analysis across multiple disciplines. IGI Global, pp 1846–1870
- Keramafar A, Amirkhani H, Bidgoly AJ (2022) Modeling tweet dependencies with graph convolutional networks for sentiment analysis. *Cognit Comput*. <https://doi.org/10.1007/s12559-021-09986-8>
- Khan MT, Ma Y, Kim J-j (2016) Term ranker: a graph-based re-ranking approach. In: The twenty-ninth international flairs conference. AAAI
- Kharde V, Sonawane P et al (2016) Sentiment analysis of twitter data: a survey of techniques. *Int J Comput Appl* 975:0975–8887
- Kim S-M, Hovy E (2004) Determining the sentiment of opinions. In: COLING 2004: Proceedings of the 20th international conference on computational linguistics. ACL, pp 1367–1373
- Kim K, Lee J (2014) Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recogn* 47(2):758–768
- Kim H, Howland P, Park H, Christianini N (2005) Dimension reduction in text classification with support vector machines. *J Mach Learn Res* 6(1):37–53
- Kim J, Kim H-U, Adamowski J, Hatami S, Jeong H (2022) Comparative study of term-weighting schemes for environmental big data using machine learning. *Environ Model Softw* 157:105536
- Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907)
- Kiritchenko S, Zhu X, Mohammad SM (2014a) Sentiment analysis of short informal texts. *J Artif Intell Res* 50:723–762
- Kiritchenko S, Zhu X, Cherry C, Mohammad S (2014b) Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). ACL, pp 437–442
- Kolkur S, Dantal G, Mahe R (2015) Study of different levels for sentiment analysis. *Int J Curr Eng Technol* 5(2):768–770
- König AC, Brill E (2006) Reducing the human overhead in text categorization. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 598–603
- Kontopoulos E, Berberidis C, Dergiades T, Bassiliades N (2013) Ontology-based sentiment analysis of twitter posts. *Expert Syst Appl* 40(10):4065–4074



- Kummer O, Savoy J (2012) Feature weighting strategies in sentiment analysis. In: SDAD 2012: the first international workshop on sentiment discovery from affective data, pp 48–55
- Kwon K, Choi C-H, Lee J, Jeong J, Cho W-S (2015) A graph based representative keywords extraction model from news articles. In: Proceedings of the 2015 international conference on big data applications and services. ACM, pp 30–36
- Lahiri S, Choudhury SR, Caragea C (2014) Keyword and keyphrase extraction using centrality measures on collocation networks. arXiv preprint [arXiv:1401.6571](https://arxiv.org/abs/1401.6571)
- Lan M, Tan CL, Su J, Lu Y (2008) Supervised and traditional term weighting methods for automatic text categorization. *IEEE Trans Pattern Anal Mach Intell* 31(4):721–735
- Le B, Nguyen H (2015) Twitter sentiment analysis using machine learning techniques. In: Advanced computational methods for knowledge engineering. Springer, pp 279–289
- Li Y-M, Li T-Y (2013) Deriving market intelligence from microblogs. *Decis Support Syst* 55(1):206–217
- Li H, Lu W (2017) Learning latent sentiment scopes for entity-level sentiment analysis. In: Thirty-first AAAI conference on artificial intelligence. AAAI, pp 3482–3489
- Li N, Wu DD (2010) Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decis Support Syst* 48(2):354–368
- Li S, Zhang H, Xu W, Chen G, Guo J (2010) Exploiting combined multi-level model for document sentiment analysis. In: 2010 20th international conference on pattern recognition. IEEE, pp 4141–4144
- Li S-K, Guan Z, Tang L-Y, Chen Z (2012) Exploiting consumer reviews for product feature ranking. *J Comput Sci Technol* 27(3):635–649
- Li X, Xie H, Chen L, Wang J, Deng X (2014) News impact on stock price return via sentiment analysis. *Knowl-Based Syst* 69:14–23
- Li S, Zhou L, Li Y (2015a) Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures. *Inf Process Manag* 51(1):58–67
- Li X, Li J, Wu Y (2015b) A global optimization approach to multi-polarity sentiment analysis. *PLoS ONE* 10(4):0124672
- Li Y, Pan Q, Yang T, Wang S, Tang J, Cambria E (2017) Learning word representations for sentiment analysis. *Cognit Comput* 9(6):843–851
- Li W, Shao W, Ji S, Cambria E (2022) Bieru: bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing* 467:73–82
- Liang B, Su H, Gui L, Cambria E, Xu R (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl-Based Syst* 235:107643
- Litvak M, Last M, Aizenman H, Gobits I, Kandel A (2011) Degext—a language-independent graph-based keyphrase extractor. In: Advances in intelligent web mastering—3 vol 86. Springer, pp 121–130
- Liu B, Zhang L (2012) A survey of opinion mining and sentiment analysis. In: Mining text data. Springer, pp 415–463
- Liu Y, Loh HT, Sun A (2009) Imbalanced text classification: a term weighting approach. *Expert Syst Appl* 36(1):690–701
- Liu H, He J, Wang T, Song W, Du X (2013) Combining user preferences and user opinions for accurate recommendation. *Electron Commer Res Appl* 12(1):14–23
- Loper E, Bird S (2002) Nltk: The natural language toolkit. In: Proceedings of the ACL-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics, vol 1. ACM, pp 63–70
- Loughran T, McDonald B (2015) The use of word lists in textual analysis. *J Behav Financ* 16(1):1–11
- Lu Q, Zhu Z, Zhang G, Kang S, Liu P (2021) Aspect-gated graph convolutional networks for aspect-based sentiment analysis. *Appl Intell* 51(7):4408–4419
- Lu Q, Sun X, Sutcliffe R, Xing Y, Zhang H (2022) Sentiment interaction and multi-graph perception with graph convolutional networks for aspect-based sentiment analysis. *Knowl-Based Syst* 256:109840
- Luo F, Li C, Cao Z (2016) Affective-feature-based sentiment analysis using svm classifier. In: 2016 IEEE 20th international conference on computer supported cooperative work in design (CSCWD). IEEE, pp 276–281
- Ma Y, Song R, Gu X, Shen Q, Xu H (2022) Multiple graph convolutional networks for aspect-based sentiment analysis. *Appl Intell* 1–14
- Majumder N, Poria S, Hazarika D, Mihalcea R, Gelbukh A, Cambria E (2019) Dialoguernn: An attentive RNN for emotion detection in conversations. In: Proceedings of the AAAI conference on artificial intelligence, vol 33. IEEE, pp 6818–6825
- Malliaros FD, Skianis K (2015) Graph-based term weighting for text categorization. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015. ACM, pp 1473–1479

- Malo P, Sinha A, Korhonen P, Wallenius J, Takala P (2014) Good debt or bad debt: detecting semantic orientations in economic texts. *J Am Soc Inf Sci* 65(4):782–796
- Manning CD, Surdeanu M, Bauer J, Finkel JR, Bethard S, McClosky D (2014) The stanford corenlp natural language processing toolkit. In: Proceedings of 52nd annual meeting of the Association for Computational Linguistics: system demonstrations. ACL, pp 55–60
- Mäntylä MV, Graziotin D, Kuutila M (2018) The evolution of sentiment analysis-a review of research topics, venues, and top cited papers. *Comput Sci Rev* 27:16–32
- Mao R, Liu Q, He K, Li W, Cambria E (2022) The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection. *IEEE Trans Affect Comput*
- Mars A, Gouider MS (2017) Big data analysis to features opinions extraction of customer. *Procedia Comput Sci* 112:906–916
- Mathew L, Bindu V (2020) A review of natural language processing techniques for sentiment analysis using pre-trained models. In: 2020 fourth international conference on computing methodologies and communication (ICCMC). IEEE, pp 340–345
- McAuley J, Leskovec J (2013) Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM conference on recommender systems. ACM, pp 165–172
- Medelyan O, Witten IH (2006) Thesaurus based automatic keyphrase indexing. In: Proceedings of the 6th ACM/IEEE-CS joint conference on digital libraries. ACM, pp 296–297
- Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J* 5(4):1093–1113
- Mehta P, Pandya S (2020) A review on sentiment analysis methodologies, practices and applications. *Int J Sci Technol Res* 9(2):601–609
- Miller GA, Beckwith R, Fellbaum C, Gross D, Miller KJ (1990) Introduction to wordnet: an on-line lexical database. *Int J Lexicogr* 3(4):235–244
- Mohammad S (2012) # emotional tweets. In: \* SEM 2012: The first joint conference on lexical and computational semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the sixth international workshop on semantic evaluation (SemEval 2012). ACL, pp 246–255
- Mohammad SM, Kiritchenko S (2015) Using hashtags to capture fine emotion categories from tweets. *Comput Intell* 31(2):301–326
- Mohammad S, Turney P (2010) Emotions evoked by common words and phrases: Using mechanical Turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text. ACL, pp 26–34
- Mohammad SM, Turney PD (2013) Crowdsourcing a word-emotion association lexicon. *Comput Intell* 29(3):436–465
- Mohammad S, Dunne C, Dorr B (2009) Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: Proceedings of the 2009 conference on empirical methods in natural language processing. ACL, pp 599–608
- Mohammad SM, Kiritchenko S, Zhu X (2013) Nrc-canada: building the state-of-the-art in sentiment analysis of tweets. In: Second joint conference on lexical and computational semantics (\* SEM), Volume 2: Proceedings of the seventh international workshop on semantic evaluation (SemEval 2013), vol 2. ACL, pp 321–327
- Moreno-Ortiz A, Fernández-Cruz J, Hernández CPC (2020) Design and evaluation of Sentiecon: a fine-grained economic/financial sentiment lexicon from a corpus of business news. In: Proceedings of The 12th language resources and evaluation conference. ACL, pp 5065–5072
- Mostafa MM (2013) More than words: social networks' text mining for consumer brand sentiments. *Expert Syst Appl* 40(10):4241–4251
- Mothe J, Ramiaandrisoa F, Rasolomanana M (2018) Automatic keyphrase extraction using graph-based methods. In: Proceedings of the 33rd annual ACM symposium on applied computing. ACM, pp 728–730
- Mullen T, Collier N (2004) Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of the 2004 conference on empirical methods in natural language processing. ACL, pp 412–418
- Nagarajan R, Nair S, Aruna P, Puviarasan N (2016) Keyword extraction using graph based approach. *Int J Adv Res Comput Sci Softw Eng* 6(10):25–29
- Narayanan R, Liu B, Choudhary A (2009) Sentiment analysis of conditional sentences. In: Proceedings of the 2009 conference on empirical methods in natural language processing. ACL and AFNLP, pp 180–189



- Nasar Z, Jaffry SW, Malik MK (2019) Textual keyword extraction and summarization: state-of-the-art. *Inf Process Manage* 56(6):102088
- Nguyen TD, Kan M-Y (2007) Keyphrase extraction in scientific publications. In: *International conference on Asian digital libraries*. Springer, pp 317–326
- Nguyen H, Nguyen M-L (2017) A deep neural architecture for sentence-level sentiment classification in twitter social networking. In: *International conference of the Pacific Association for Computational Linguistics*. Springer, pp 15–27
- Nielsen FÅ (2011) A new anew: evaluation of a word list for sentiment analysis in microblogs. In: *1st workshop on making sense of Microposts: big things come in small packages*, pp 93–98
- Nielsen FÅ (2017) *afinn project*
- O’Keefe T, Koprinska I (2009) Feature selection and weighting methods in sentiment analysis. In: *Proceedings of the 14th Australasian document computing symposium*, Sydney, pp 67–74
- Oliveira N, Cortez P, Areal N (2014) Automatic creation of stock market lexicons for sentiment analysis using stocktwits data. In: *Proceedings of the 18th international database engineering & applications symposium*. ACM, pp 115–123
- Ouyang Y, Guo B, Zhang J, Yu Z, Zhou X (2017) Sentistory: multi-grained sentiment analysis and event summarization with crowdsourced social media data. *Pers Ubiquit Comput* 21(1):97–111
- Pang B, Lee L (2004) A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. *ACM*
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing*, Morristown, NJ, USA. Association for Computational Linguistics, pp 79–86
- Passi K, Motisariya J (2022) Twitter sentiment analysis of the 2019 indian election. In: *IOT with smart systems*. Springer, pp 805–814
- Patil P, Yalagi P (2016) Sentiment analysis levels and techniques: a survey. *Int J Innov Eng Technol* 6:523
- Patil G, Galande V, Kekan V, Dange K (2014) Sentiment analysis using support vector machine. *Int J Innov Res Comput Commun Eng* 2(1):2607–2612
- Pavitha N, Pungliya V, Raut A, Bhonsle R, Purohit A, Patel A, Shashidhar R (2022) Movie recommendation and sentiment analysis using machine learning. In: *Global transitions proceedings*. Elsevier
- Pitogo VA, Ramos CDL (2020) Social media enabled e-participation: a lexicon-based sentiment analysis using unsupervised machine learning. In: *Proceedings of the 13th international conference on theory and practice of electronic governance*, pp 518–528. *ACM*
- Poria S, Gelbukh A, Hussain A, Howard N, Das D, Bandyopadhyay S (2013) Enhanced senticnet with affective labels for concept-based opinion mining. *IEEE Intell Syst* 28(2):31–38
- Poria S, Cambria E, Hazarika D, Majumder N, Zadeh A, Morency L-P (2017) Context-dependent sentiment analysis in user-generated videos. In: *Proceedings of the 55th annual meeting of the Association for Computational Linguistics (volume 1: Long Papers)*. *ACL*, pp 873–883
- Prasad AG, Sanjana S, Bhat SM, Harish B (2017) Sentiment analysis for sarcasm detection on streaming short text data. In: *2017 2nd International conference on knowledge engineering and applications (ICKEA)*. *IEEE*, pp 1–5
- Prastyo PH, Sumi AS, Dian AW, Permanasari AE (2020) Tweets responding to the Indonesian government’s handling of covid-19: sentiment analysis using svm with normalized poly kernel. *J Inf Syst Eng Bus Intell* 6(2):112–122
- Priyadarshini I, Cotton C (2021) A novel lstm-cnn-grid search-based deep neural network for sentiment analysis. *J Supercomput* 77(12):13911–13932
- Quan C, Ren F (2014) Unsupervised product feature extraction for feature-oriented opinion determination. *Inf Sci* 272:16–28
- Rabelo JC, Prudêncio RB, Barros FA (2012) Using link structure to infer opinions in social networks. In: *2012 IEEE international conference on systems, man, and cybernetics (SMC)*. *IEEE*, pp 681–685
- Rajput A (2020) Natural language processing, sentiment analysis, and clinical analytics. In: *Lytras MD, Sarirete A (eds) Innovation in health informatics*. Elsevier, Academic Press, pp 79–97
- Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowl-Based Syst* 89:14–46
- Ravinuthala MKV et al (2016) Thematic text graph: a text representation technique for keyword weighting in extractive summarization system. *Int J Inf Eng Electron Bus* 8(4)
- Read J, Carroll J (2009) Weakly supervised techniques for domain-independent sentiment classification. In: *Proceedings of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion*. *ACM*, pp 45–52
- Ren F, Sohrab MG (2013) Class-indexing-based term weighting for automatic text classification. *Inf Sci* 236:109–125

- Ren R, Wu DD, Liu T (2018) Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Syst J* 13(1):760–770
- Reyes A, Rosso P (2012) Making objective decisions from subjective data: detecting irony in customer reviews. *Decis Support Syst* 53(4):754–760
- Rui H, Liu Y, Whinston A (2013) Whose and what chatter matters? The effect of tweets on movie sales. *Decis Support Syst* 55(4):863–870
- Saif H, Fernández M, He Y, Alani H (2014) On stopwords, filtering and data sparsity for sentiment analysis of Twitter, pp 810–817
- Salari N, Shohaimi S, Najafi F, Nallappan M, Karishnarajah I (2014) A novel hybrid classification model of genetic algorithms, modified k-nearest neighbor and developed backpropagation neural network. *PLoS ONE* 9(11):112987
- Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Inf Process Manage* 24(5):513–523
- Santos G, Mota VF, Benevenuto F, Silva TH (2020) Neutrality may matter: sentiment analysis in reviews of AIRBNB, booking, and Couchsurfing in Brazil and USA. *Soc Netw Anal Min* 10(1):1–13
- Sarzynska-Wawer J, Wawer A, Pawlak A, Szymanowska J, Stefaniak I, Jarkiewicz M, Okruszek L (2021) Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Res* 304:114135
- Saxena A, Reddy H, Saxena P (2022) Introduction to sentiment analysis covering basics, tools, evaluation metrics, challenges, and applications. In: *Principles of social networking*. Springer, pp 249–277
- Schouten K, Frasincar F (2015) Survey on aspect-level sentiment analysis. *IEEE Trans Knowl Data Eng* 28(3):813–830
- Sebastiani F, Debole F (2003) Supervised term weighting for automated text categorization. In: *Proceeding the 18th ACM symposium on applied computing*. ACM, pp 784–788
- Serrano-Guerrero J, Olivas JA, Romero FP, Herrera-Viedma E (2015) Sentiment analysis: a review and comparative analysis of web services. *Inf Sci* 311:18–38
- Sharma A, Dey S (2012) A comparative study of feature selection and machine learning techniques for sentiment analysis. In: *Proceedings of the 2012 ACM research in applied computation symposium*, pp 1–7
- Shi W, Zheng W, Yu JX, Cheng H, Zou L (2017) Keyphrase extraction using knowledge graphs. *Data Sci Eng* 2(4):275–288
- Shimada K, Hashimoto D, Endo T (2009) A graph-based approach for sentiment sentence extraction. In: *Pacific-Asia conference on knowledge discovery and data mining*. Springer, pp 38–48
- Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez N, Velásquez F, Díaz-Rangel I, Suárez-Guerra S, Trevino A, Gordon J (2013) Empirical study of machine learning based approach for opinion mining in tweets. In: *Mexican international conference on artificial intelligence*. Springer, pp 1–14
- Smith H, Cipolli W (2022) The Instagram/Facebook ban on graphic self-harm imagery: a sentiment analysis and topic modeling approach. *Policy Internet* 14(1):170–185
- Sokolova M, Lapalme G (2007) Performance measures in classification of human communications. In: *Conference of the Canadian Society for computational studies of intelligence*. Springer, pp 159–170
- Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manage* 45(4):427–437
- Solangi YA, Solangi ZA, Aarain S, Abro A, Mallah GA, Shah A (2018) Review on natural language processing (NLP) and its toolkits for opinion mining and sentiment analysis. In: *2018 IEEE 5th international conference on engineering technologies and applied sciences (ICETAS)*. IEEE, pp 1–4
- Soubraylu S, Rajalakshmi R (2021) Hybrid convolutional bidirectional recurrent neural network based sentiment analysis on movie reviews. *Comput Intell* 37(2):735–757
- Sousa MG, Sakiyama K, de Souza Rodrigues L, Moraes PH, Fernandes ER, Matsubara ET (2019) Bert for stock market sentiment analysis. In: *2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI)*. IEEE, pp 1597–1601
- Stagner R (1940) The cross-out technique as a method in public opinion analysis. *J Soc Psychol* 11(1):79–90
- Staiano J, Guerini M (2014) Depechemood: a lexicon for emotion analysis from crowd-annotated news. *ACL*, pp 427–433. arXiv preprint [arXiv:1405.1605](https://arxiv.org/abs/1405.1605)
- Stone PJ, Hunt EB (1963) A computer approach to content analysis: studies using the general inquirer system. In: *Proceedings of the May 21–23, 1963, Spring Joint Computer Conference*. ACL, pp 241–256
- Subrahmanian VS, Reforgiato D (2008) Ava: adjective-verb-adverb combinations for sentiment analysis. *IEEE Intell Syst* 23(4):43–50

- Taboada M (2016) Sentiment analysis: an overview from linguistics. *Ann Rev Linguist* 2:325–347. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- Tamilselvam S, Nagar S, Mishra A, Dey K (2017) Graph based sentiment aggregation using conceptnet ontology. In: *Proceedings of the eighth international joint conference on natural language processing (Volume 1: Long Papers)*, vol 1. ACL, pp 525–535
- Tan S, Zhang J (2008) An empirical study of sentiment analysis for Chinese documents. *Expert Syst Appl* 34(4):2622–2629
- Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P (2011) User-level sentiment analysis incorporating social networks. In: *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery*, pp 1397–1405
- Tan LK-W, Na J-C, Theng Y-L, Chang K (2012) Phrase-level sentiment polarity classification using rule-based typed dependencies and additional complex phrases consideration. *J Comput Sci Technol* 27(3):650–666
- Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M (2015) Sentiment embeddings with applications to sentiment analysis. *IEEE Trans Knowl Data Eng* 28(2):496–509
- Tembhurne JV, Diwan T (2021) Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks. *Multimedia Tools Appl* 80(5):6871–6910
- Thelwall M, Buckley K (2013) Topic-based sentiment analysis for the social web: the role of mood and issue-related words. *J Am Soc Inform Sci Technol* 64(8):1608–1617
- Theng Y-L (2004) Design and usability of digital libraries: case studies in the Asia Pacific: case studies in the Asia Pacific. IGI Global, pp 129–152
- Tian Y, Chen G, Song Y (2021) Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In: *Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies. ACL*, pp 2910–2922
- Titov I, McDonald R (2008) A joint model of text and aspect ratings for sentiment summarization. In: *Proceedings of ACL-08: HLT. ACL*, pp 308–316
- Trinh S, Nguyen L, Vo M (2018) Combining lexicon-based and learning-based methods for sentiment analysis for product reviews in Vietnamese language. In: *International conference on computer and information science. Springer*, pp 57–75
- Tripathy A, Agrawal A, Rath SK (2015) Classification of sentimental reviews using machine learning techniques. *Procedia Comput Sci* 57:821–829
- Tsai AC-R, Wu C-E, Tsai RT-H, Hsu JY (2013) Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Intell Syst* 28(2):22–30
- Turney PD (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp 417–424
- Vakali A, Chatzakou D, Koutsonikola V, Andreadis G (2013) Social data sentiment analysis in smart environments-extending dual polarities for crowd pulse capturing. In: *International conference on data management technologies and applications*, vol 2. SCITEPRESS, pp 175–182
- Valakunde N, Patwardhan M (2013) Multi-aspect and multi-class based document sentiment analysis of educational data catering accreditation process. In: *2013 International conference on cloud & ubiquitous computing & emerging technologies. IEEE*, pp 188–192
- Valdivia A, Luzión MV, Herrera F (2017) Neutrality in the sentiment analysis problem based on fuzzy majority. In: *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE*, pp 1–6
- Valdivia A, Luzón MV, Cambria E, Herrera F (2018) Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Inf Fusion* 44:126–135
- Vega-Oliveros DA, Gomes PS, Milios EE, Berton L (2019) A multi-centrality index for graph-based keyword extraction. *Inf Process Manag* 56(6):102063
- Verma S (2022) Sentiment analysis of public services for smart society: literature review and future research directions. *Gov Inf Q* 39:101708
- Vidanagama D, Silva A, Karunananda A (2022) Ontology based sentiment analysis for fake review detection. *Expert Syst Appl* 206:117869
- Wakade S, Shekar C, Liszka KJ, Chan C-C (2012) Text mining for sentiment analysis of twitter data. In: *Proceedings of the international conference on information and knowledge engineering (IKE). The Steering Committee of The World Congress in Computer Science, Computer ...*, pp 1–6
- Wang Z, Joo V, Tong C, Xin X, Chin HC (2014) Anomaly detection through enhanced sentiment analysis on social media data. In: *2014 IEEE 6th international conference on cloud computing technology and science. IEEE*, pp 917–922

- Wang T, Cai Y, Leung H-f, Cai Z, Min H (2015) Entropy-based term weighting schemes for text categorization in VSM. In: 2015 IEEE 27th international conference on tools with artificial intelligence (ICTAI). IEEE, pp 325–332
- Wang J, Li C, Xia C (2018) Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Appl Math Comput* 334:388–400
- Wang Z, Ho S-B, Cambria E (2020) Multi-level fine-scaled sentiment sensing with ambivalence handling. *Int J Uncertain Fuzz Knowl-Based Syst* 28(04):683–697
- Wang X, Li F, Zhang Z, Xu G, Zhang J, Sun X (2021) A unified position-aware convolutional neural network for aspect based sentiment analysis. *Neurocomputing* 450:91–103
- Wang J, Zhang Y, Yu L-C, Zhang X (2022a) Contextual sentiment embeddings via bi-directional GRU language model. *Knowl-Based Syst* 235:107663
- Wang Z, Gao P, Chu X (2022b) Sentiment analysis from customer-generated online videos on product review using topic modeling and multi-attention BLSTM. *Adv Eng Inform* 52:101588
- Wankhade M, Rao ACS, Kulkarni C (2022) A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev* 1–50
- Wawre SV, Deshmukh SN (2016) Sentiment classification using machine learning techniques. *IJSR* 5(4):819–821
- Wiebe JM (1990) Recognizing subjective sentences: a computational investigation of narrative text. State University of New York at Buffalo
- Wiebe J, Mihalcea R (2006) Word sense and subjectivity. In: Proceedings of the 21st International conference on computational linguistics and 44th annual meeting of the Association for Computational Linguistics, pp 1065–1072
- Wilson T, Wiebe J (2003) Annotating opinions in the world press. In: Proceedings of the fourth SIGdial workshop of discourse and dialogue. ACL, pp 13–22
- Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of human language technology conference and conference on empirical methods in natural language processing. ACL, pp 347–354
- Wu Y, Zhang Q, Huang X-J, Wu L (2011) Structural opinion mining for graph-based sentiment representation. In: Proceedings of the 2011 conference on empirical methods in natural language processing. ACL, pp 1332–1341
- Xia R, Zong C, Li S (2011) Ensemble of feature sets and classification algorithms for sentiment classification. *Inf Sci* 181(6):1138–1152
- Yadav A, Vishwakarma DK (2020) Sentiment analysis using deep learning architectures: a review. *Artif Intell Rev* 53(6):4335–4385
- Yadav CS, Sharan A, Joshi ML (2014) Semantic graph based approach for text mining. In: 2014 international conference on issues and challenges in intelligent computing techniques (ICICT). IEEE, pp 596–601
- Yan X, Huang T (2015) Tibetan sentence sentiment analysis based on the maximum entropy model. In: 2015 10th international conference on broadband and wireless computing, communication and applications (BWCCA). IEEE, pp 594–597
- Yan Z, Xing M, Zhang D, Ma B (2015) Exprs: an extended pagerank method for product feature extraction from online consumer reviews. *Inf Manag* 52(7):850–858
- Yavari A, Hassanpour H, Rahimpour Cami B, Mahdavi M (2022) Election prediction based on sentiment analysis using twitter data. *Int J Eng* 35(2):372–379
- Ye Q, Zhang Z, Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Syst Appl* 36(3):6527–6535
- Yenter A, Verma A (2017) Deep cnn-lstm with combined kernels from multiple branches for imdb review sentiment analysis. In: 2017 IEEE 8th annual ubiquitous computing, electronics and mobile communication conference (UEMCON). IEEE, pp 540–546
- Yu B, Zhang S (2022) A novel weight-oriented graph convolutional network for aspect-based sentiment analysis. *J Supercomput* 1–26
- Yu X, Liu Y, Huang X, An A (2010) Mining online reviews for predicting sales performance: a case study in the movie domain. *IEEE Trans Knowl Data Eng* 24(4):720–734
- Yu Y, Duan W, Cao Q (2013) The impact of social and conventional media on firm equity value: a sentiment analysis approach. *Decis Support Syst* 55(4):919–926
- Zad S, Heidari M, Jones JH, Uzuner O (2021) A survey on concept-level sentiment analysis techniques of textual data. In: 2021 IEEE World AI IoT Congress (AIIoT). IEEE, pp 0285–0291
- Zainuddin N, Selamat A (2014) Sentiment analysis using support vector machine. In: 2014 international conference on computer, communications, and control technology (I4CT), pp 333–337. IEEE

- Zhan J, Loh HT, Liu Y (2009) Gather customer concerns from online product reviews-a text summarization approach. *Expert Syst Appl* 36(2):2107–2115
- Zhang K, Xu H, Tang J, Li J (2006) Keyword extraction using support vector machine. In: International conference on web-age information management. Springer, pp 85–96
- Zhang L, Ghosh R, Dekhil M, Hsu M, Liu B (2011) Combining lexicon-based and learning-based methods for twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011 89, 1–8. HP Laboratories
- Zhang W, Xu H, Wan W (2012) Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Syst Appl* 39(11):10283–10291
- Zhang H, Gan W, Jiang B (2014) Machine learning and lexicon based methods for sentiment classification: a survey. In: 2014 11th web information system and application conference. IEEE, pp 262–265
- Zhang Y, Zhou Y, Yao J (2020) Feature extraction with tf-idf and game-theoretic shadowed sets. In: International conference on information processing and management of uncertainty in knowledge-based systems, pp 722–733. Springer
- Zhang Q, Yi GY, Chen L-P, He W (2021) Text mining and sentiment analysis of covid-19 tweets. *arXiv preprint arXiv:2106.15354*
- Zhang K, Zhang K, Zhang M, Zhao H, Liu Q, Wu W, Chen E (2022) Incorporating dynamic semantics into pre-trained language model for aspect-based sentiment analysis. In: Findings of the Association for Computational Linguistics: ACL 2022. ACL, pp 3599–3610
- Zhao WX, Jiang J, He J, Song Y, Achanauparp P, Lim E-P, Li X (2011) Topical keyphrase extraction from twitter. In: Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies, pp 379–388
- Zhao Z, Tang M, Tang W, Wang C, Chen X (2022) Graph convolutional network with multiple weight mechanisms for aspect-based sentiment analysis. *Neurocomputing* 500:124–134
- Zhou J, Tian J, Wang R, Wu Y, Xiao W, He L (2020) Sentix: a sentiment-aware pre-trained model for cross-domain sentiment analysis. In: Proceedings of the 28th international conference on computational linguistics. ACL, pp 568–579
- Zhu X, Kiritchenko S, Mohammad S (2014) Nrc-canada-2014: recent improvements in the sentiment analysis of tweets. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). ACL, pp 443–447
- Zhu X, Zhu L, Guo J, Liang S, Dietze S (2021) Gl-gcn: Global and local dependency guided graph convolutional networks for aspect-based sentiment classification. *Expert Syst Appl* 186:115712
- Zhu H, Zheng Z, Soleymani M, Nevatia R (2022) Self-supervised learning for sentiment analysis via image-text matching. In: ICASSP 2022–2022 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1710–1714
- Zin HM, Mustapha N, Murad MAA, Sharef NM (2017) The effects of pre-processing strategies in sentiment analysis of online movie reviews. In: AIP conference proceedings, vol 1891. AIP Publishing LLC, p. 020089

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.