


An Analysis of Sentiment: Methods, Applications, and Challenges [†]

Harish Dutt Sharma *  and Parul Goyal

Department of Computer Application, School of CA&IT, Shri Guru Ram Rai University,
Dehradun 248001, Uttarakhand, India; profdrparul@gmail.com

* Correspondence: harish.research106@gmail.com

[†] Presented at the International Conference on Recent Advances on Science and Engineering, Dubai, United Arab Emirates, 4–5 October 2023.

Abstract: Sentiment analysis involves contextually examining text to identify and extract subjective information from source material. It aids businesses in comprehending the public sentiment surrounding their brand, product, or service while monitoring online discussions. Nevertheless, analyzing social media content is often limited to basic sentiment analysis and simple count-based metrics. Devices that allow the collection of huge amounts of unstructured, opinionated data are becoming increasingly connected with humans. Everyday-activity-related comments and evaluations have been obtained as a result of the advances in Internet-based services like social media platforms and blogs. This study supplies a comprehensive assessment of sentiment analysis approaches to provide academics with a global perspective on the analysis of feelings and its associated domain, applications, and challenges. To comprehend the applications of sentiment analysis, this article provides a detailed explanation of the technique for performing this activity. To comprehend the benefits and drawbacks of each method, they are then evaluated, compared, and discussed. To establish future perspectives, the difficulties of sentiment analysis are finally evaluated.

Keywords: sentiment analysis; text mining; social media; emotion detection; classification



Citation: Sharma, H.D.; Goyal, P. An Analysis of Sentiment: Methods, Applications, and Challenges. *Eng. Proc.* **2023**, *59*, 68. <https://doi.org/10.3390/engproc2023059068>

Academic Editors: Nithesh Naik,
Rajiv Selvam, Pavan Hiremath,
Suhas Kowshik CS and Ritesh
Ramakrishna Bhat

Published: 19 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The computational investigation of human behavior, perspectives, and feelings towards an item is referred to as “sentiment analysis (SA) or opinions mining (OM)”. A person, an event, or a topic may be represented by the object. The Natural Language Processing (NLP) approach known as SA tries to extract opinions and attitudes from texts [1,2]. The objectives of SA are to identify if a particular text reflects a positive or negative attitude; hence, this task may be thought of as a text-classification issue [3]. Nevertheless, while sentiment analysis may appear to be a simple procedure, it involves the evaluation of other NLP subtasks, including sarcasm and subjectivity identification [4,5].

Nowadays, sentiment analysis is widely considered by businesses, governments, and organizations in addition to researchers [6]. Despite the long and distinguished history of using public opinion in decision making, there should be several prior publications acknowledging SA, a branch of NLP [7]. Even though SA has evolved in the modern era, it is still functional. The issues, approaches, applications, and algorithms used in sentiment analysis will all be examined in this paper. It provides the task with simple tables, flowcharts, and graphs that demonstrate comparative data analysis. According to our knowledge, existing surveys typically overlook certain SA methods in aid of lexicon-based, machine learning (ML), and hybrid techniques. The goals of this study are to develop the conceptual framework and discuss its applications and challenges in this sector. The key contributions of the survey may be summed up as follows:

- The definition of the sentiment analysis procedure in detail, as well as the identification of well-known techniques for performing it in this paper; it was achieved via the reviews of a variety of publications.
- Analyses of the different techniques to select which is best for a certain application.
- To comprehend more easily obtainable techniques like ML, lexicon-based models, and hybrid models, we categorize and describe commonly used SA methodologies.

Section 2, phases of SA, is how the literature review is structured. Section 3 explains the “Data Gathering, characteristic extracting, and method of Feature Selection”, including each step from data extraction through to the different sentiment analysis techniques. The basic sentiment analysis approach and its summaries are detailed in Section 4. Sections 5 and 6 contain applications and challenges of sentiment analysis in several fields, and finally, Section 7 concludes the work.

2. Phases of Sentiment Analysis

At various stages, sentiment analysis has been researched. Thoughts and views can, nevertheless, primarily be examined during the document phase, sentence phase, or aspect phase [8,9]. The stages of sentiment analysis are depicted in Figure 1. The first two phases are both interesting and quite difficult. The third step, however, is more challenging, since it conducts a fine-grained study. Here is a quick explanation of each stage:

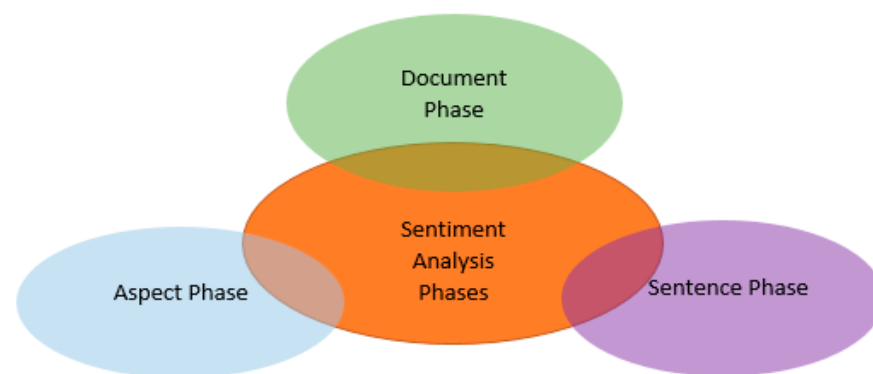


Figure 1. Various phases of sentiment analysis.

2.1. Document Phase

A single polarity is assigned to the whole document after conducting SA at the document phase. Not many organizations use SA of this kind. The classification of a book’s chapters or pages as good, negative, or neutral could be performed using this method. At this stage, the document can be categorized using “supervised and unsupervised learning techniques” [10]. The most important difficulties in document-level sentiment analysis are “cross-domain” and “cross-linguistic” sentiment analysis [11].

2.2. Sentence Phase

At this stage of analysis, each sentence is analyzed to identify its polarity. This is extremely powerful when the subject is associated with a wide range of emotions [12]. This phase of categorization is connected to subjective classification [13]. For some applications, the data-level analytics view is not enough [14], and prior research on sentence-level analysis aimed at identifying subjective sentences is needed. Nevertheless, there are more challenging assignments, including coping with conditional words or unclear statements [15].

2.3. Aspect Phase

This phase conducts a fine-grained analysis to identify attitudes towards certain aspects of things. As a result, the complexities at this level assist in understanding precisely what people like and dislike [16]. According to [17], the fundamental job of sentiment

analysis is aspect extraction. This can be accomplished using either implicit or explicit characteristics. Mowlaei et al. [18] suggested a method for aspect-based SA employing adaptive aspect-based lexicons. To achieve a higher performance, two or more levels could be integrated to complete sentiment analysis rather than just one level. Mai and Le [19] suggested a combined strategy of sentiment analysis at the aspect and phrase layers for product comments on YouTube.

3. Sentiment Analysis Pre-Processing and Feature Extraction

Figure 2 illustrates the general sentiment analysis process. Data have been collected, having been retrieved in different formats from a wide range of sources, transformed into text, and then analyzed with NLP methods. The processing step especially includes features like “text pre-processing, feature extraction, and feature selection”.

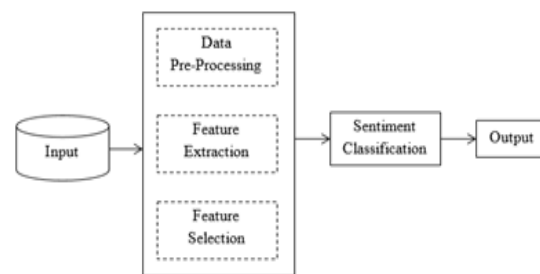


Figure 2. General flow of sentiment analysis.

3.1. Data Pre-Processing

Specifically with social media, there is a plethora of unstructured data produced. These data may be noisy in their raw form and include a variety of syntax and grammatical problems [20]. The overall process includes numerous typical tasks:

- Tokenization;
- Stop-Word Removal;
- Part of Speech Tagging;
- Lemmatization.

3.2. Feature Extraction

In sentiment classification, feature extraction (FE) is a crucial operation since it requires sifting through text data to find important information that may be used to improve the model’s performance. Due to the difficulty in extracting features from the text, Venugopalan and Gupta (2015) [21] also included features in their research.

3.2.1. Terms Frequency (TF)

It is among the easiest methods for expressing characteristics that are increasingly often being utilized in sentiment analysis and other NLP applications, such as information retrieval. According to Sharma et al. (2013) [22], it aids in determining one word, or a uni-gram, or a group of between two and three terms, which could be a bi-gram or tri-gram.

3.2.2. Parts of Speech (PoS) Tag

Grammatical tagging is the technique of categorizing words in a corpus of texts (corpus) according to their syntax and semantics. “Tokens are divided into nouns, verbs, pronouns, adverbs, adjectives, and prepositions”. PoS taggers, that are included in NLTK or Spacy, may be used for this job. Stanford PoS-tagger is most frequently used in the research [23].

3.2.3. Negations

These are the phrases that could transform the meaning of a sentence and modify the polarity of an opinion. Negation terms like not, cannot, neither, never, nowhere, none, etc.,

are often used. Excluding all negative phrases from stop words might increase processing expenses and decrease the model's accuracy since not every word in the sentence will cause the polarity to shift. Negative words should be used with great care [24].

3.2.4. Bag of Words (BoW)

BoW is among the most basic ways for deleting text features. BoW would describe the word consequences in a document. The repertoire of words needed to create a vector for every statement is represented by a bag. The primary issue with this approach is that it disregards the grammatical significance of the text. TF-IDF is used to analyze the performance of the BoW method, which generally makes it more effective.

3.3. Feature Selection (FS)

An attribute of the data is described by a feature, as was previously explained. A feature may be redundant, relevant, or neither. Several feature selection (FS) approaches are used to eliminate duplicate and unnecessary characteristics. To lower the size of the feature dimension space and increase the precision of emotion classification, FS is a technique that involves finding and removing extraneous and unnecessary characteristics from the feature list [25].

Lexicon-based approaches and statistical techniques are used in feature selection [26]. Features in lexicon-based techniques are developed by humans. To develop a smaller feature set, the procedure is often begun by gathering phrases with strong feeling. The next step is to add more terms to this collection using web resources or synonym detection. The Senti-WordNet lexicon is a well-known illustration of this methodology. The four standard classifications for statistical techniques are the filter method, wrapper approach, embedding method, and hybrid model [27].

3.3.1. Filter Method

This method of feature selection is the most used one. Based on the overall features of the training data, it chooses features absent from the use of other machine learning techniques. Various statistical criteria are used to rate the feature, and the features with the highest rankings are then selected [28].

3.3.2. Wrapper Method

This method is dependent on ML approaches since it depends on the results of the algorithm. As a consequence of such dependence, approaches are usually continuous and computationally time-consuming, even though they can discover the optimum feature set for a certain prediction method.

3.3.3. Embedded Method

This framework incorporates the FS into the operation of the analysis approach. It employs classification algorithms with feature selection functionality built in. However, the method is algorithm specific [29].

3.3.4. Hybrid Method

In this technique, filter and wrapper approaches are combined; hybrid techniques typically use different methods to create the ideal feature subset. For sentiment analysis, several hybrid feature selection algorithms have been created [30].

4. Classification of Sentiment Analysis

The study of sentiment is a fast-growing and dynamic research field with numerous applications. The purpose is to enhance the analysis of sentiment performance and address challenges related to this topic. Based on different viewpoints, it is important to integrate the current techniques into sentiment analysis. Nevertheless, the majority of research often separates SA methodologies into three groups: ML methodologies, lexicon-based

methodologies, and hybrid methodologies [31]. The lexicon-based technique makes use of the emotion lexicon, which is a collection of expressions that are often used to describe either good or negative emotions [32]. “A hybrid technique, on the other hand, combines machine learning and lexicon-based methods to enhance sentiment analysis performance”. Figure 3 depicts the structure of the SA techniques.

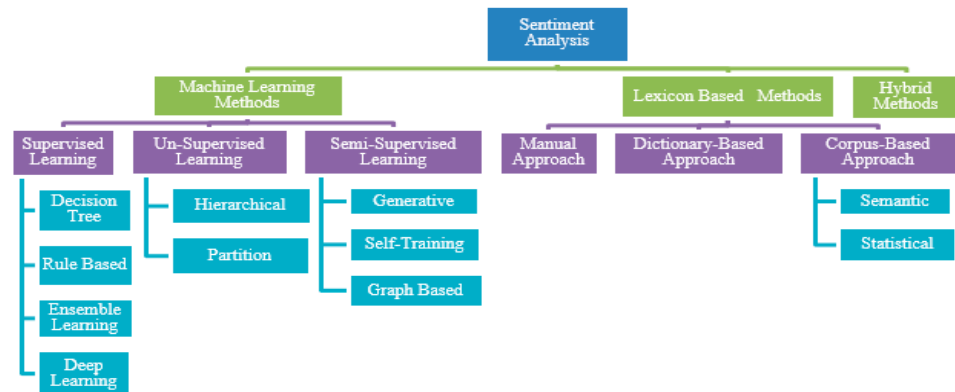


Figure 3. Classification of sentiment analysis.

4.1. Machine Learning Approaches

Based on train and test datasets, ML techniques are used to categorize sentiment polarity. According to [33], these methods may be categorized as follows: “unsupervised-learning”, “semi-supervised-learning”, and “supervised learning”. When a classification program contains a predetermined set of classes, the supervised technique is utilized; however, when identifying this set is difficult due to a lack of labeled data, the unsupervised method may be critical. In contrast, the semi-supervised approach can be applied to unlabeled datasets that contain some labeled instances. To aid the agent’s integration with its environment and maximize cumulative rewards, reinforcement learning algorithms utilize trial-and-error techniques.

However, classifiers trained on specific data do not perform well for different domains [34].

These algorithms must be trained before being applied to real data. Data in text format can be used to extract features. SA systems can be trained to go beyond simple content to detect informational content, profanity, and misuse of words. Commonly used algorithms include the following:

4.1.1. Naive Bayes (NB)

Both categorization and training are completed using the NB approach. The Bayes rule-based Bayesian classification method NB is used. The Bayes rule is used by NB, a probabilistic classifier, to forecast the likelihood that a specific collection of features will be included with a particular label. When the proportion of data used for training is small, NB is commonly utilized. Positive NB categorization was 10% more accurate than negative NB categorization. As an outcome of its implementation, average accuracy was enhanced.

4.1.2. Support Vector Machine (SVM)

The SVM approach is used in that technique to analyze data and set decision boundaries, which makes use of hyper-planes. SVMs are a kind of supervised learning method that is not probabilistic and is widely used for classification problems. Finding the hyperplane that divides the data into subgroups most effectively is the main goal of SVM. Because of this, SVM aims for the hyperplane with the maximum achievable margin. Linear SVM and VADER were used by Borg and Boldt (2020) [35] to predict the sentiment of customer reviews.

The linear SVM classifier was implemented enormously [36], with an F1 score of 83.4% and a mean AUC of 0.896. Moreover, their algorithm showed an improvement in

email exchanges that was predicted using the emotion of unseen emails. In contrast to the conventional binary classification, the problem emphasized that the subjectivity of opinion and legitimacy of the expresser must be evaluated [37]. To integrate perspectives from microblogs, a methodology (Wu et al., 2020) [38] was suggested. The issues stated in the views that were relevant to the users' requests were located and extracted, and the opinions were then classified using SVM. For the experiment, ref. [39] was also implemented using Twitter tweet data. They considered that compiling the opinions for microblogs would be advantageous.

4.1.3. Machine Learning Approaches

Multiplying the value of an input through a weight value is a step in the machine-learning technique called logistic regression. It is the classifier, which generates knowledge of the input characteristics that aid in classifying between positive and negative classes. Any category of independent variables is possible. According to the LR technique [40], the dependent variable is binary and the predicting variables demonstrate little to no multicollinearity.

4.1.4. K-Nearest Neighbors (KNN)

Although KNN is not frequently used in sentiment analysis, the KNN technique has been found to yield good results if properly trained. The K value can be chosen using any hyper-parameter tuning process, such as grid search or randomized search cross-validation. The polarity can be determined via soft addition or hard voting depending on the scores of the K nearest neighbors.

4.1.5. Decision Tree (DT)

A tree is formed using the training example in the supervised learning method known as the DT Classifier to categorize the text's polarity. These are employed to identify the trait that best captures reality and ought to be at the base. Yan-Yan et al. (2010) proposed a graph-based reporting strategy to combine sentence-level and sentence-level data [41]. Tagged information that may be used to identify between legitimate and fraudulent reviews is presented in the work of [42].

4.2. *Lexicon-Based Approach*

Lexicons are arrays of tokens that have a predetermined score that indicates whether the content is neutral, positive, or negative. The Lexicon-based approach adds positive, negative, and neutral values to each token for a certain review or text. Finally, based on the highest value of the individual evaluations, the statement is assigned an overall polarity. As a result, the text is separated into single-word tokens, and each token's polarity is computed and averaged. The lexicon-based method is especially useful for feature and sentence-level sentiment analysis. Because no training data is necessary, this strategy may be classified as unsupervised.

4.2.1. Corpus-Based Approach

The model uses the semantic and syntactic structure to determine the meaning of a sentence and a syntactic model or another similar model is used for describing the concept of tokens and their orientation in the larger body after using a list of words and their orientation. This approach is unique and works well for learning from multiple domains. The following methods are known for corpus-based methods. Statistical methods and semantic methods are described below.

Statistical method: statistical methods can be used to identify facts or association patterns between ideas.

Semantic method: This method calculates the similarity score of the markers for emotional evaluation. Widely used for this purpose is WordNet.

4.2.2. Dictionary-Based Method

This strategy is predicated on the idea that whereas antonyms and synonymous terms have opposing emotional polarity, respectively, the well-known dictionaries are used to generate the sentiment lexicons in this method. Like other dictionary-based methods, the main problem with dictionary-based concepts is the inability to define conceptual words with specific expressions, making them unsuitable for content and specific distribution. Additionally, it is difficult and time consuming to write dependency code.

4.3. Hybrid Methods

Lexicon and machine learning techniques are combined in the hybrid model. To process ambiguity and include the context of sentiment words, it integrates the efficiency of lexical analysis with the adaptability of ML methodologies [43]. The key benefit of the hybrid technique is that it inherits both stability from the lexicon-based model and good accuracy from ML. The hybrid system integrates techniques from the two preceding methods to overcome their disadvantages and enhance their advantages.

5. Comparison of Different Classification Techniques

The accuracy, precision, and recall of different classification techniques based on machine learning techniques are evaluated and compared in this section. Table 1 shows the results of the study.

Table 1. Evaluation results of various machine learning techniques.

Classification Techniques	Accuracy	Precision	Recall
Naïve Bayes	86.01%	79.25%	81.26%
SVM	94.05%	88.27%	87.34%
LR	90.23%	80.61%	79.02%
KNN	89.47%	85.19%	86.58%
DT	91.53%	88.86%	87.19%

In the comparison of machine learning techniques, the SVM has achieved the highest accuracy of 94.05%, whereas the Decision Tree (DT) has achieved the highest precision rate of 88.86% when compared to all machine techniques, and the SVM has achieved the highest recall rate of 87.34% as compared to all other techniques. Figure 4 shows the graphical representation of the overall comparison of classification techniques of sentiment analysis.

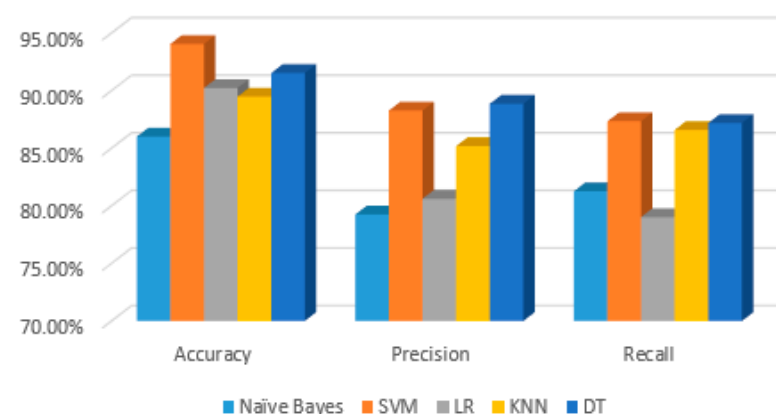


Figure 4. Performance comparison of classification techniques.

6. Applications of Sentiment Analysis

The following list of relevant sectors and areas where sentiment analysis is used is provided:

6.1. Business Analysis

SA has several advantages in the context of business intelligence (BI). Businesses may also use sentiment analysis data to enhance goods, look into customer comments, and build creative marketing plans.

6.2. Health Care and Medical Domain

Recently, sentiment analysis has been used in this industry, among others. Surveys, Twitter [44], blogs, news stories, reviews, and other sources may all be used to gather data. The usage of sentiment analysis and other NLP applications is being actively researched by domain specialists [45] (Ebadi et al., 2021).

6.3. Review Analysis

The field of entertainment makes substantial use of sentiment analysis. Reviews of movies, television series, and short films can be examined to establish the opinion of the public [46] (Kumar et al., 2019). The travel industry is working to improve the customer experience through machine learning and web-based customer insights, often focused on smart, data-driven decision making.

6.4. Customer Voice

Compiling and evaluating all user input from call centers, emails, surveys, chats, and the web, data may be categorized and organized using sentiment analysis to find patterns and recurrent problems and concerns.

6.5. Social Media Monitoring

Social media data tracks client opinions in real time, around the clock, seven days a week, allowing businesses to respond promptly to unfavorable comments and better their reputation when they receive positive comments.

7. Challenges in Sentiment Analysis

SA primarily focuses on analyzing evaluations and comments about various people and processing them to extract any helpful information. The SA process is influenced by several factors, all of which must be managed properly to produce the final classification or clustering report. Some of these difficulties are described further below.

7.1. Sarcasm Handling

Sarcasm is often referred to as the use of words that convey the opposite of what is intended, e.g., "What a good batsman he is, he scores zero in every other inning". These phrases are challenging to identify.

7.2. Domain Dependency

In SA, words are mostly implemented as an analytical element. However, the meaning of the terms varies from sentence to sentence. There are not many terms whose definitions vary depending on the context. Additionally, there are terms known as contronyms that have contradictory meanings in certain contexts. Identifying the context in which a term is used might be difficult since it determines how the text is analyzed and, ultimately, how the result turns out.

7.3. Negations

The inclusion of negative words in the content can drastically change the precise meaning of the phrase in which they appear. As a result, when reading reviews, it is critical to keep these phrases in mind.

7.4. Spam Detection

Reviews are examined as part of sentiment analysis. However, a small amount of qualitative analysis has been performed so far to determine whether the evaluations are genuine or were provided by someone who is indeed a valid reviewer. The service is reviewed favorably or unfavorably by a large number of people who do not know the company's product or service. Knowing which reviews are false and which are real is quite tough to do, but it is crucial for SA.

8. Conclusions

A brief description of the analysis of sentiment and associated techniques is provided in this publication. The major goal of this article is to explore and categorize the most prevalent sentiment analysis categorization methods. Following a brief introduction of crucial steps including data collection and FS, various layers of sentiment analysis were first reviewed. The classification process then moved on to sentiment categorization approaches. Then, several strategies for categorizing sentiments were constructed. In this field, supervised ML algorithms are frequently used due to their ease of use and high precision.

Author Contributions: H.D.S. and P.G. conceptualized the study and helped with its design. Each author revised the work and gave their approval to the finished product. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Acknowledgments: Parul Goyal, who served as my research advisor, has my sincere gratitude for his outstanding support and direction.

Conflicts of Interest: Regarding the research, writing, or publication of this paper, the authors declare no conflict of interest.

References

1. Birjali, M.; Kasri, M.; Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl. Based Syst.* **2021**, *226*, 107134. [\[CrossRef\]](#)
2. Chaturvedi, I.; Cambria, E.; Welsch, R.E.; Herrera, F. Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Inf. Fusion* **2018**, *44*, 65–77. [\[CrossRef\]](#)
3. Choi, Y.; Lee, H. Data properties and the performance of sentiment classification for electronic commerce applications. *Inf. Syst. Front.* **2017**, *19*, 993–1012. [\[CrossRef\]](#)
4. Cambria, E.; Das, D.; Bandyopadhyay, S.; Feraco, A. (Eds.) *A Practical Guide to Sentiment Analysis*; Springer: Cham, Switzerland, 2017.
5. Valdivia, A.; Luzón, M.V.; Cambria, E.; Herrera, F. Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Inf. Fusion* **2018**, *44*, 126–135. [\[CrossRef\]](#)
6. Sánchez-Rada, J.F.; Iglesias, C.A. Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Inf. Fusion* **2019**, *52*, 344–356. [\[CrossRef\]](#)
7. Wankhade, M.; Rao, A.C.S.; Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* **2022**, *55*, 5731–5780. [\[CrossRef\]](#)
8. Do, H.H.; Prasad, P.W.C.; Maag, A.; Alsadoon, A. Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Syst. Appl.* **2019**, *118*, 272–299. [\[CrossRef\]](#)
9. Aggarwal, C.C. *Machine Learning for Text*; Springer: Cham, Switzerland, 2018; Volume 848.
10. Bhatia, P.; Ji, Y.; Eisenstein, J. Better document-level sentiment analysis from rst discourse parsing. *arXiv* **2015**, arXiv:1509.01599.
11. Saunders, D. Domain Adaptation for Neural Machine Translation. Doctoral Dissertation, University of Cambridge, Cambridge, UK, 2021.
12. Yang, B.; Cardie, C. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 22–27 June 2014; Volume 1, pp. 325–335.

13. Rao, G.; Huang, W.; Feng, Z.; Cong, Q. LSTM with sentence representations for document-level sentiment classification. *Neurocomputing* **2018**, *308*, 49–57. [\[CrossRef\]](#)
14. Behdenna, S.; Barigou, F.; Belalem, G. Document level sentiment analysis: A survey. *EAI Endorsed Trans. Context-Aware Syst. Appl.* **2018**, *4*, e2. [\[CrossRef\]](#)
15. Ferrari, A.; Esuli, A. An NLP approach for cross-domain ambiguity detection in requirements engineering. *Autom. Softw. Eng.* **2019**, *26*, 559–598. [\[CrossRef\]](#)
16. Indurkha, N.; Damerau, F.J. *Handbook of Natural Language Processing*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2010.
17. Tubishat, M.; Idris, N.; Abushariah, M.A. Implicit aspect extraction in sentiment analysis: Review, taxonomy, opportunities, and open challenges. *Inf. Process. Manag.* **2018**, *54*, 545–563. [\[CrossRef\]](#)
18. Mowlaei, M.E.; Abadeh, M.S.; Keshavarz, H. Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Syst. Appl.* **2020**, *148*, 113234. [\[CrossRef\]](#)
19. Mai, L.; Le, B. Joint sentence and aspect-level sentiment analysis of product comments. *Ann. Oper. Res.* **2021**, *300*, 493–513. [\[CrossRef\]](#)
20. Liu, B. *Sentiment Analysis and Opinion Mining*; Synthesis Lectures on Human Language Technologies; Springer: Cham, Switzerland, 2012; Volume 5, pp. 1–167.
21. Venugopalan, M.; Gupta, D. Exploring sentiment analysis on twitter data. In Proceedings of the 2015 Eighth International Conference on Contemporary Computing (IC3), Noida, India, 20–22 August 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 241–247.
22. Sharma, A.; Lyons, J.; Dehzangi, A.; Paliwal, K.K. A feature extraction technique using bi-gram probabilities of position specific scoring matrix for protein fold recognition. *J. Theor. Biol.* **2013**, *320*, 41–46. [\[CrossRef\]](#)
23. Weerasooriya, T.; Perera, N.; Liyanage, S.R. A method to extract essential keywords from a tweet using NLP tools. In Proceedings of the 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), Negombo, Sri Lanka, 1–3 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 29–34.
24. George, D.R.; Rovniak, L.S.; Kraschnewski, J.L. Dangers and opportunities for social media in medicine. *Clin. Obstet. Gynecol.* **2013**, *56*, 453–462. [\[CrossRef\]](#)
25. Ahmad, S.R.; Bakar, A.A.; Yaakub, M.R. A review of feature selection techniques in sentiment analysis. *Intell. Data Anal.* **2019**, *23*, 159–189. [\[CrossRef\]](#)
26. Kumar, R.; Kaur, J. Random forest-based sarcastic tweet classification using multiple feature collection. In *Multimedia Big Data Computing for IoT Applications*; Springer: Singapore, 2020; pp. 131–160.
27. Hoque, N.; Bhattacharyya, D.K.; Kalita, J.K. MIFS-ND: A mutual information-based feature selection method. *Expert Syst. Appl.* **2014**, *41*, 6371–6385. [\[CrossRef\]](#)
28. Adomavicius, G.; Kwon, Y. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Trans. Knowl. Data Eng.* **2011**, *24*, 896–911. [\[CrossRef\]](#)
29. Das, H.; Naik, B.; Behera, H.S. A Jaya algorithm based wrapper method for optimal feature selection in supervised classification. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 3851–3863. [\[CrossRef\]](#)
30. Chiew, K.L.; Tan, C.L.; Wong, K.; Yong, K.S.; Tiong, W.K. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Inf. Sci.* **2019**, *484*, 153–166. [\[CrossRef\]](#)
31. Sankar, H.; Subramaniaswamy, V. Investigating sentiment analysis using machine learning approach. In Proceedings of the 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 7–8 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 87–92.
32. Jurek, A.; Mulvenna, M.D.; Bi, Y. Improved lexicon-based sentiment analysis for social media analytics. *Secur. Inform.* **2015**, *4*, 1–13. [\[CrossRef\]](#)
33. Yusof, N.N.; Mohamed, A.; Abdul-Rahman, S. Reviewing classification approaches in sentiment analysis. In Proceedings of the International Conference on Soft Computing in Data Science, Putrajaya, Malaysia, 2–3 September 2015; Springer: Singapore, 2015; pp. 43–53.
34. Yoo, G.; Nam, J. A hybrid approach to sentiment analysis enhanced by sentiment lexicons and polarity shifting devices. In Proceedings of the 13th Workshop on Asian Language Resources, Miyazaki, Japan, 7 May 2018; pp. 21–28.
35. Borg, A.; Boldt, M. Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Syst. Appl.* **2020**, *162*, 113746. [\[CrossRef\]](#)
36. Li, F.; Wang, W.; Xu, J.; Yi, J.; Wang, Q. Comparative study on vulnerability assessment for urban buried gas pipeline network based on SVM and ANN methods. *Process Saf. Environ. Prot.* **2019**, *122*, 23–32. [\[CrossRef\]](#)
37. Xia, H.; Yang, Y.; Pan, X.; Zhang, Z.; An, W. Sentiment analysis for online reviews using conditional random fields and support vector machines. *Electron. Commer. Res.* **2020**, *20*, 343–360. [\[CrossRef\]](#)
38. Wu, P.; Li, X.; Shen, S.; He, D. Social media opinion summarization using emotion cognition and convolutional neural networks. *Int. J. Inf. Manag.* **2020**, *51*, 101978. [\[CrossRef\]](#)
39. Ali, S.M.; Noorian, Z.; Bagheri, E.; Ding, C.; Al-Obeidat, F. Topic and sentiment aware microblog summarization for twitter. *J. Intell. Inf. Syst.* **2020**, *54*, 129–156. [\[CrossRef\]](#)
40. Hamdan, H.; Bellot, P.; Bechet, F. Lsislif: Crf and logistic regression for opinion target extraction and sentiment polarity analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 753–758.

41. Zhao, Y.-Y.; Qin, B.; Liu, T. Integrating intra-and inter-document evidences for improving sentence sentiment classification. *Acta Autom. Sin.* **2010**, *36*, 1417–1425. [[CrossRef](#)]
42. Jain, P.K.; Pamula, R.; Ansari, S. A supervised machine learning approach for the credibility assessment of user-generated content. *Wirel. Pers. Commun.* **2021**, *118*, 2469–2485. [[CrossRef](#)]
43. Gupta, I.; Joshi, N. Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic. *J. Intell. Syst.* **2020**, *29*, 1611–1625. [[CrossRef](#)]
44. Carvalho, J.; Plastino, A. On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis. *Artif. Intell. Rev.* **2021**, *54*, 1887–1936. [[CrossRef](#)]
45. Ebadi, A.; Xi, P.; Tremblay, S.; Spencer, B.; Pall, R.; Wong, A. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. *Scientometrics* **2021**, *126*, 725–739. [[CrossRef](#)]
46. Kumar, S.; Yadava, M.; Roy, P.P. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Inf. Fusion* **2019**, *52*, 41–52. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.