

Review

Not peer-reviewed version

A Systematic Review of Sentiment Analysis Systems Applied to Textual Data

[Phuong Dao Quoc](#) * and Vuong M. Ngo

Posted Date: 3 October 2025

doi: 10.20944/preprints202510.0194.v1

Keywords: domain; application; dataset; task; algorithm




Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Review

A Systematic Review of Sentiment Analysis Systems Applied to Textual Data

Vuong M. Ngo  and Phuong Q. Dao * 

Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam

* Correspondence: phuongdq.24Ai@ou.edu.vn

Abstract

With the explosion of social networks, the amount of data produced by online users every day is huge. Sentiment Analysis (SA) serves as a powerful tool for extracting opinions and emotions from text, offering valuable insights to support decision-making across various domains. Developed by computer scientists, SA is applied in product and marketing analysis, strategy monitoring, financial and political forecasting, misinformation detection, health, education, and public opinion analysis. While recent SA studies focus on models and techniques, there remains a need for more efficient, purpose-specific models. SA performance depends on data quality, feature selection, domain context, and classifier choice. We employ the PRISMA methodology to narrow our search scope, focusing on 4,709 papers published by reputable sources such as Springer, Elsevier, IEEE, ACM, and others from 2020 to present. Through an iterative review of titles, abstracts, and full texts, we selected 34 papers relevant to our topic, including 30 research papers and 4 survey papers, for in-depth analysis. Our survey offers a comprehensive view of developments in SA, covering domains, applications, datasets, techniques, and future directions. It highlights knowledge gaps and suggests potential for multidisciplinary use and future research.

Keywords: domain; application; dataset; task; algorithm

1. Introduction

1.1. Background and Motivation

The explosion of user-generated data is driven by advancements in internet technology and the rapid growth of social media platforms [1]. This data—ranging from text and images to videos and sounds—is vast, diverse, and originates from various structured and unstructured sources. Analyzing such data is crucial for big data applications [2,3]. Techniques such as Machine Learning (ML), Deep Learning (DL), Opinion Mining, and Text Mining have long been used to support this analysis, although challenges remain due to data volume and variability [4–7]. Social platforms such as Twitter, YouTube, Facebook, and blogs generate content in multiple formats, often lacking proper grammar or structure. The abundance of user opinions shared on these platforms has led to increased interest in SA, as reflected by the growing number of publications [1,8,9].

Organizations now process large amounts of text data—emails, chat transcripts, social media comments, and reviews. SA tools classify sentiment (positive, negative, or neutral) to extract insights into human emotions and guide decision-making. In marketing, SA gauges consumer sentiment to improve strategies. In politics, it helps forecast trends and assess public opinion. In finance, SA supports investment decisions by analyzing market sentiment.

Moreover, SA plays a role in company strategy monitoring, detecting fake news and hate speech, and promoting responsible information sharing. It also supports healthcare by identifying trends in patient feedback and public health discussions. In education, SA helps personalize learning through analysis of student sentiment. Additionally, it provides valuable insights for public opinion analysis in policymaking and research.

Given these applications, there is a growing need for further research to develop efficient SA models tailored to specific domains. A comprehensive survey can help identify research gaps and analyze key factors such as data quality, feature selection, domain context, and algorithm choice that influence model performance.

1.2. Related Work

Recent surveys [1,9–11] highlight the promise of ML, DL, and hybrid techniques in automating SA, with supervised methods favored for their simplicity and high accuracy. Rodriguez-Ibanez et al. [1] reviewed SA in social networks, noting marketing as the most studied domain, followed by politics, economics, and health—driven by abundant data from Twitter and commercial sites. Ezhilarasan et al. [11] examined DL for product review SA, while Bordoloi et al. [10] stressed the need for domain-specific keyword extraction to avoid errors from generalized dictionaries. [9] used Naive Bayes and SVM as benchmarks for evaluating new SA models.

However, many surveys lack a cross-domain view of SA datasets, models, and techniques. [11] focused mainly on product reviews and rule-based methods; [1,9] did not assess the impact of domain-specific dataset properties on accuracy; and [10] covered domains and techniques but lacked detail on applied models or future directions. In contrast, our survey provides a comprehensive review of SA domains, datasets, and techniques, based on 30 carefully selected studies from over 4,709 papers published between 2020 and 2024, focusing on ML, DL, and hybrid models to offer a current and wide-ranging perspective.

1.3. Contribution

By exploring a diverse array of domains, tasks, methodologies, datasets, and the challenges encountered in current research papers, this paper aims to offer insights into the evolving landscape of SA. The contributions are outlined as follows:

1. We offer a thorough examination of domains, datasets, and tasks specifically within the field of SA. Additionally, we provide insightful analyses derived from the compiled information.
2. Our exploration both ML and DL algorithms, including Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), played crucial roles. Subsequently, we present a comprehensive analysis of the advantages and disadvantages associated with these approaches in the context of SA.
3. We delve into the challenges confronting SA models, addressing issues such as the dynamic nature of language, context-dependent interpretations, and the prospect of constructing a knowledge graph representation for semantic analytics or establishing sentiment scores for each entity.
4. We draw from the findings in the SA research publication, we present and succinctly outline the essential steps involved in constructing a SA system.

2. Review Methodology

We focus on reviewing papers employing ML or DL models for SA systems. Our overall methodology, which is used to identify the relevant papers, is based on PRISMA¹ [12,13]. The following criteria for research paper selection were used:

- Published by Springer, Elsevier, IEEE or ACM which respectively have libraries being Springer Nature², Science Direct³, IEEE Xplore⁴ or ACM DL⁵.
- Published from 01/01/2020 to 31/12/2024.
- Written in English, regardless of geographical region or dataset language.

¹ <https://www.prisma-statement.org/Extensions/Protocols>

² <https://link.springer.com/>

³ <https://www.sciencedirect.com>

⁴ <https://ieeexplore.ieee.org>

⁵ <https://dl.acm.org>

- Paper title, keywords, or abstract must include: ("Sentiment Analysis" OR "Opinion Mining") AND ("Machine Learning" OR "Deep Learning" OR Classification) AND (Datasets OR Applications), adapted to each library's Boolean search format.

Ethical considerations together with relevance to the research topics, are crucial. These include ethical standards in research conduct, participant consent, data handling, and other pertinent ethical considerations. However, it's important to clarify that the ethical assessment of the literature review falls outside the scope of our paper. As shown in Table 1, with search criteria above, there are 4,709 downloaded papers through the advanced search functions of the publisher libraries. After reviewing the title and abstract of the papers, we identified 206 relevant papers. The precise count of papers for each publisher and year after the "downloading" and "reviewing title and abstract" phases is depicted in in Figure 1. We observed a year-on-year increase in the number of papers.

For each of these 206 papers, we continued to review the full text carefully, and selected 32 papers which are most relevant to the purpose of our paper. To avoid missing other relevant papers, we searched on Google Scholar and did not limit the publisher or publication year of papers. This search led us to find an extra 2 relevant papers, resulting in a total of 34 relevant papers. This final set contained 30 research papers and 4 review papers in SA applications and methods. We presented and compared the 4 review papers [1,9–11], in Section 1. The subsequent sections will involve a detailed analysis, classification, and discussion of the 30 research papers.

Table 1. The number of related literatures in publishers.

Publisher	The number of literatures		
	with search criteria	After downloading	After reviewing title and abstract
Springer	1,886	64	9 (Res ¹ : 8, Sur ² : 1)
Elsevier	2,066	91	11 (Res: 10, Sur: 1)
IEEE	465	39	9 (Res: 9, Sur: 0)
ACM	292	12	3 (Res: 3, Sur: 0)
Total	4,709	206	32 (Res: 30, Sur: 2)
Others with unlimited publisher and pub. year			2 (Res: 0, Sur: 2)
Final Review			34 (Res: 30, Sur: 4)

¹ Res: the number of research papers.

² Sur: the number of survey papers.

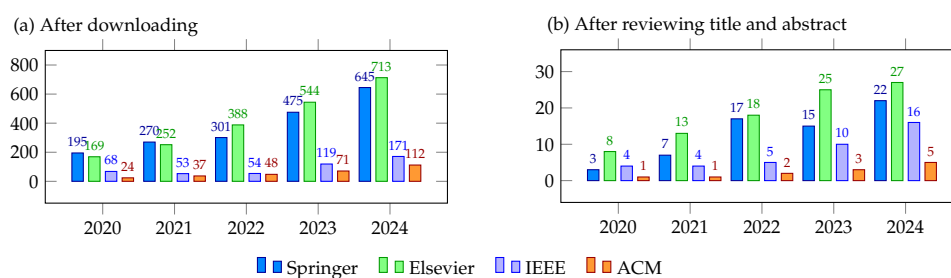


Figure 1. Volume of literature by year.

3. Classification and Analysis

3.1. Overview

The 30 current research papers are classified and analysed in Tables 2. These papers are presented in an order organised by publishers and categorized based on the domains of SA, the datasets used, the addressed SA tasks, and the applied models. Each paper is categorized under the respective publisher's name, and the year of publication is specified. The domain and task of each research paper are also outlined. Lastly, the proposed models and their corresponding accuracies are provided.

Moreover, these papers are organised by domain. Each paper is categorized in detail including its datasets, algorithms in models and developments. The datasets provide information about their source, name and features. These categories in Tables 2 is analysed in the detail below.

Table 2. Models, Domains, Dataset, Tasks and Algorithm in the research.

No	Models	Year	Domains	Dataset ¹	Task ²	Algorithm and Accuracy
Springer						
1	DL [14]	2020	Management	Self+NO	impr	HAN, Acc: 0.80
2	FE (DL)+DL [15]	2021	Finance	Self+NO	new-m	IKN-ConvLSTM, Acc: 0.98
3	Corpus+DL [16]	2022	Detection	Self+NO	new-m	T5-Sentence+CNN, F1: 0.75
4	ML [17]	2022	Education	Self+NO	comp	SVM, Acc: 0.63
5	DL [18]	2023	Finance	Self+O	new-m	CrudeBERT, Acc: 0.98
6	DL [19]	2024	Health	3rd+O	impr	CT-BERT, Acc: 0.87
7	FE (DL)+DL [20]	2024	Politics	Self+O	new-m	CNN+LSTM, Acc: 0.92
8	Corpus+DL [21]	2024	Technology	Self+NO	new-m	FastText+BiLSTM, F1: 0.71
Elsevier						
1	FE (DL)+DL [22]	2020	Health	3rd+O	comp	BERT+LSTM, F1: 0.947
2	Corpus+ML [23]	2020	Education	Self+NO	expe	TF-IDF+RF, Acc: 0.96
3	Corpus+ML [24]	2023	Product	3rd+O	impr	TF-IDF+SVM, Acc: 0.99
4	FE (DL)+DL [25]	2023	Detection	Self+O	expe	fine-tuning XLNet, Acc: 0.95
5	FE (DL)+DL [26]	2023	Health	3rd+O	expe	CNN+LSTM, Acc: 0.94
6	FE (DL)+DL [27]	2023	Product	3rd+O	new-m	B-MLCNN, Acc: 0.95
7	Corpus+ML [28]	2024	Finance	Self+NO	impr	TF-IDF+SVM, Acc: 0.82
8	ML [29]	2024	Politics	3rd+O	expe	SVM, Acc: 0.72
9	Corpus+DL [30]	2024	Product	3rd+O	new-m	GloVe+BiGRU, Acc: 0.98
10	DL [31]	2024	Product	3rd+O	comp	DistilBERT, Acc: 0.96
IEEE						
1	FE (DL)+DL [32]	2020	Education	Self+NO	new-m	Multi-AFM, Acc: 0.946
2	DL [33]	2020	Finance	3rd+O	expe	BART-Large, Acc: 0.947
3	FE (DL)+DL [34]	2023	Detection	3rd+O	new-m	CNN-BiLSTM+BERT, Acc: 0.94
4	DL [35]	2023	Product	3rd+O	impr	KGAN+RoBERTa, Acc: 0.9438
5	FE (DL)+DL [36]	2024	Finance	3rd+O	impr	MET-GAT, Acc: 0.8556
6	FE (DL)+DL [37]	2024	Finance	3rd+O	new-m	FinBERT-BiGRU, Acc: 0.95
7	FE (DL)+DL [38]	2024	Product	3rd+O	new-m	ESSGCN, Acc: 0.8786
8	FE (DL)+DL [39]	2024	Product	3rd+O	new-m	enBiLSTM-GRU, Acc: 0.94
9	FE (DL)+DL [40]	2024	Product	3rd+O	new-m	BERT-BiGRU, Acc: 0.94
ACM						
1	FE (DL)+ML [41]	2023	Product	Self+NO	expe	ConvNN+SVM, Acc: 0.845
2	FE (DL)+ML [42]	2023	Education	Self+NO	new-m	CNN,LSTM+SVM, Acc: 0.93
3	Corpus+DL [43]	2024	Health	Self+NO	new-m	TextInception+BiGRU, Acc: 0.99

¹ Self = self-built; 3rd = third-party; NO = no Open for access; O = Open for access

² new-m = new model; comp = comparison; impr = improvement; expe = experiment

3.2. Domains

Derived from an extensive dataset, the information shared by users offers significant potential for analyzing opinions across diverse domains, including financial predictions, product and marketing analysis, and health analysis, as presented in Tables 2.

In recent years, text sentiment analysis and opinion mining have gained popularity in the finance [15,18,28,33,36,37] due to their ability to convey the thoughts and feelings of corporate stakeholders such as managers and investors, including textual emotion in the forecasting of firm financial performance resulting in a significant improvement in performance. Furthermore, in product and marketing analysis using social platforms, consumers can share their opinions and experiences. These platforms provide for product-related comments, which is a useful tool for any decision-making process. [24,27,30,31,35,38,39,41]. In the field of health, research involves examining public opinions in order to provide in making decisions with a better understanding of how the public views the disease [19,22,26,43].

Online social media platforms are widely used for communication, sharing life events, and expressing opinions on news and trending topics. However, interactions can sometimes escalate into hate speech or abusive exchanges, where SA applications help detect fake news and harmful content [16,25,34]. In the education sector, SA is applied to service analysis [17,23,32,42], supporting reputation management and educational marketing. Similarly, it aids in monitoring company strategies [14]. In

politics and forecasting [20,29], online news portals play a key role, with SA technologies providing insights into public opinion [21].

3.3. Tasks

Table 2 presents the four kinds of tasks: new model, comparison, improvement, experiment. For example, some new models are proposed in each domain such as Finance: IKN-ConvLSTM [15], CrudeBERT [18]; B-MLCNN [27]. In [15], Nti et al. proposed a new framework called IKN-ConvLSTM. The model is built based on a hybrid approach; it is a combination of CNN and LSTM to predict stock price movements. This CNN model extracted sixty-two (62) features with an accuracy of 88.75%.

In experiments in [33], Mishev et al showed that the NLP transformers had superior performances compared to other evaluated approaches; the results can be applied in finance. In [37], Yekrangi et al developed optimized embeddings for improvement (impr), producing the best results across all models, the LSTM with the fine-tuned embedding layer added as the first layer is the best model. Consequently, the combination of fine-tuned embedding and LSTM performed better than any other model in the study with an accuracy of 83.9%.

3.4. Models

Figure 2 provides a detailed overview of algorithms and the number of prior research papers within the applied model categories, including ML/DL, and hybrid approaches. Combining with information about models and algorithms in Tables 2, we analyse the model categories and their respective algorithms below.

Machine Learning: SVM, a discriminative classifier, is widely used for text classification. In a comparative study [17], four models—Naive Bayes (NB), SVM, J48DT, and RF—were evaluated in classifying students' responses. Among them, SVM showed the best performance.

Deep Learning: Transformer-based models have revolutionized NLP by effectively handling long-range dependencies and enabling parallel text processing, making them more efficient and scalable than traditional RNNs and CNNs. Their core strength lies in the attention mechanism, which allows the model to focus on relevant parts of the input while capturing contextual relationships between all words in a sentence. Leveraging this mechanism, models such as BERT, RoBERTa, and GPT have achieved state-of-the-art results in tasks including machine translation, question answering, named entity recognition, and sentiment analysis. Pre-trained on large corpora and fine-tuned for specific applications, these models excel even with limited training data, benefiting from contextual awareness, bidirectional context, transfer learning, and large-scale pre-training. Such capabilities make transformers particularly effective for sentiment analysis in domains like finance [18,33] and product reviews [35,40].

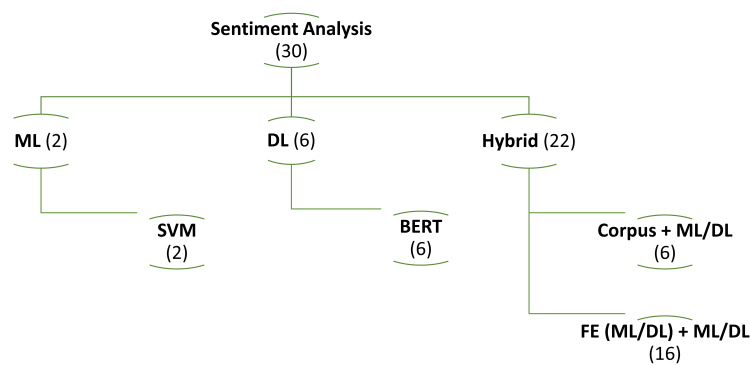


Figure 2. Models and Algorithms of Sentiment Analysis.

Hybrid Models: In addition to standalone models, hybrid approaches have shown strong potential in SA. Chakravarthi et al. [16] integrated a CNN with T5-Sentence for speech detection, while Meena et al. [26] combined CNN and LSTM to detect sentiment polarity in monkeypox tweets. Such

combinations leverage CNNs for capturing local patterns and sequence models for long-range dependencies, improving performance on long textual opinions. As shown in Table 2, hybrid methods fall into two categories: corpus + ML/DL and feature extraction (FE) (ML/DL) + ML/DL.

In the corpus + ML/DL approach, output from a corpus-based method is used as training data for sentiment classifiers built with ML/DL techniques. The FE (ML/DL) + ML/DL approach instead uses an ML/DL model as a feature extractor to generate input for subsequent ML/DL layers. In both cases, performance is evaluated on unlabeled testing data. Results indicate that combining corpus-based or ML/DL-based feature extraction with ML/DL classification yields better performance than using either approach alone.

3.5. Datasets

Table 2 serves as a comprehensive overview of two distinct types of data sources: Self (self-built by authors) and 3rd (third-party supplies). Beyond merely identifying the source, the table offers valuable insights into two critical properties of these data sources, namely NO (no Open for access) and O (Open for access). This classification provides an important context for understanding the availability and accessibility of the datasets used in the discussed SA studies.

For a more detailed illustration, consider the case of [14], located at the bottom of Table 2, where Capuano et al. extensively discussed the dataset employed for training and testing their model. This dataset, comprising over 30,000 items, was constructed with approximately 40% sourced from the customer service of Analyst Group4, showcasing the active involvement of the authors in creating a significant part of the dataset (Self). However, it's noteworthy that access to this portion of the dataset is restricted, requiring authorization due to its confidential nature (NO).

Conversely, the remaining 60% of the dataset was obtained from third-party public sources (3rd), indicating external contributions. This portion of the dataset is accessible to the public (O), highlighting a more open and transparent approach to data sourcing. This nuanced exploration of data source types and access properties enhances our understanding of the SA models discussed, underscoring the significance of transparency and accessibility in the realm of data-driven research.

3.6. Suggested Directions for Future Research

In the financial sector, [15] addressed merging six heterogeneous data sources and suggested automation with GANs or autoencoders, while [18] proposed training on full article content and extending economic models to include inflation and interest rates. For product reviews, [27,35] explored explicit polarity detection across datasets, and [38,41] tested BERT-based embeddings. In health analysis, [22] aimed to reduce reliance on annotated corpora using semantic and contextual features, [26] recommended iterative robust methods, and [36] reported gains from DL and transformers. For detection, [16] suggested larger datasets with finer classification, [25] proposed alternative datasets, and [34] highlighted multilingual text and transformer models. In education, [23] tested hybrid bi-/trigram models, [32] suggested transformer-based multi-class sentiment analysis, and large language models with clustering were noted as promising. In the political domain, dataset expansion was recommended, and in management, integrating sentiment analysis into CRM or information systems was proposed.

4. Results and Discussion

4.1. Domains

Utilizing the data categorized in Tables 2, we present a comprehensive statistical analysis depicting the evolution of sentiment analysis over recent years in Figure 3. Additionally, the application of sentiment analysis across various domains is illustrated in Figure 4. These visual representations offer valuable insights into the trends and the widespread implementation of sentiment analysis in different contexts.

Figure 3 shows the number of the research papers published increasing every year. Especially, SA is an essential research in 2023 and 2024 with approximately 73% of the selected research papers due to the rapidly growth of social medias and SA is a potential tool to discover and understand public sentiments. Figure 4 shows 8 different domains that researchers focused on in recent years, with the most popular being Product (~ 27%), Finance (~ 20%), Health (~ 17%), Education (~ 13%).

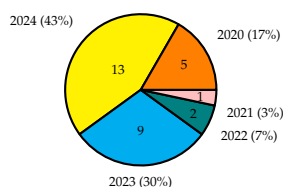


Figure 3. categorized by year.

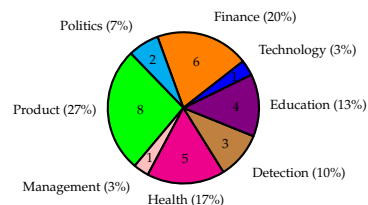


Figure 4. categorized by domain.

4.2. Tasks

Figure 5 presents the breakdown of the type of task addressed in the selected papers from the literature. The tasks addressed focused on: (1) a proposed new model; (2) make an experiment on model in other domain; (3) comparison of models; and (4) improvement an existing model by tuning hyper-parameters or adding layers. The majority (~ 50%) of the research analysed proposed new models. In terms of popularity, the remaining focused on experiment (~ 20%), improvement (~ 20%), and comparison (~ 10%). One explanation of this breakdown can be increasing hybrid approaches (~ 73%) and a proposed new model is better than previous other models, as shown in Figure 6.

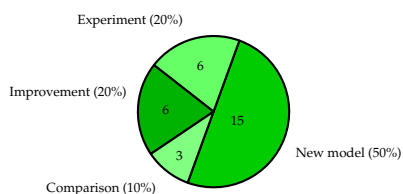


Figure 5. categorized by tasks.

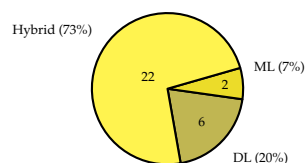


Figure 6. categorized by models & alg.

4.3. Models and Algorithms

Figure 6 presents the breakdown of the analysis and modelling techniques proposed in the selected literature. The most popular model is Hybrid (~ 73%). We also identified three major topics as: (1) noisy/imbalance nature of the data [15,18]; (2) the hybrid models used in almost domains such as Product Reviews [27,30,39], Health [22,26,43], Detection [16,25,34], Education and NLP [21,23,32,42], but DL models (~ 20%) still be used in Finance [18,33], (3) the representation of the feature vector in the corpus, ML or DL is the first step in Hybrid model. Also, based on the short analysis of the dataset, getting them labelled from the experts or linguistics is important during dataset preparation, especially in field of Detection [25]. Next, we provided an overview of various models utilized in sentiment analysis of each domain. Our examination of the surveyed articles revealed that some researchers made attempts to categorize sentiment analysis. Table 2 also shows that while both hybrid and DL models have been deployed in Product Reviews, Health, Detection, Education, Finance, and they are also used in combination with new approaches such as: Knowledge graphs [26], Semantic score together with context [22] remains more prevalent in research efforts.

4.4. Datasets

In terms of dataset categories (Figure 7), the majority (~ 53%) of the literature used datasets that were generated from a third-party such as social media and subsequently made publicly available to the community. And a smaller percentage (~ 37%) of the papers used self-generated and closed data, which is not publicly available to the research community. There are 17 open datasets created by

third-parties. In that, 7 of them are in field of product, 5 of them are in health, 2 of them are in finance, and another is technology, politics, and detection. Much smaller percentages of papers used datasets that were self-generated and then made available ($\sim 10\%$).

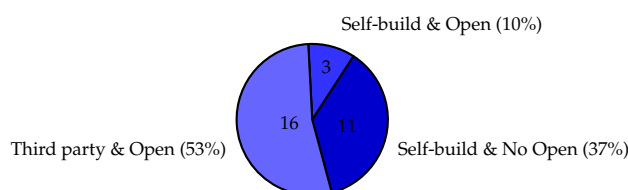


Figure 7. Prior research is classified based on the utilized datasets.

4.5. Suggested Directions for Future Research

From the survey, the financial sector faces challenges in integrating multiple data sources, requiring advances in fusion strategies, optimization, modal structures, and modality selection. Future work could leverage Knowledge Graphs to capture structural relationships in financial texts and enhance forecasting. For product reviews, a key focus is identifying explicit polarity based on text position across multiple datasets. In health analysis, priorities include reducing dependence on annotated corpora, exploring alternative knowledge embeddings, assessing transformer architectures, and combining semantic with contextual features for better fine-tuning. For detection, larger and more diverse datasets with finer classification are recommended. In education, early experiments with hybrid bi-/tri-gram models could be expanded, while in NLP technology, large language models and clustering approaches offer promising potential. In the political domain, expanding dataset size is highlighted, and in management, integrating sentiment analysis into CRM or information systems is proposed.

Recent work shows a growing adoption of transformer-based models for SA in finance [18,33], products [40], health [19], and detection [34]. By applying attention mechanisms, transformers generate contextual embeddings that capture semantic relationships, enabling state-of-the-art performance across diverse NLP tasks. They have become the preferred choice for many real-time applications; however, improvements are still needed in domain identification, domain-specific context understanding, multilingual datasets, and open-source code availability—especially for non-English languages.

5. Conclusions

With the rapid growth of social networks, vast amounts of user-generated data are produced daily. SA systems can automatically process this text to determine perspectives and provide insights into human emotions and opinions, supporting decision-making, improving outcomes, and fostering deeper understanding.

This paper has summarized, classified, and analyzed recent state-of-the-art SA research, covering domains, tasks, datasets, developments, and models. We have also outlined a generic process for developing an SA system. Traditional ML techniques such as SVM and NB laid the groundwork for modern DL models, including CNNs, RNNs, and transformer-based architectures like BERT. Hybrid models, in particular, have shown strong performance across diverse SA applications.

Looking forward, research should explore transfer learning to improve SA in low-resource languages, integrate semantic and knowledge graphs into hybrid models, and extend SA methodologies to tasks such as cyberbullying detection, depression detection, and fake news identification. Further opportunities include applying SA in manufacturing, psychiatric diagnostics, and leveraging specialized datasets within transformer-based learning.

References

1. Rodríguez-Ibáñez, M.; et al. A review on sentiment analysis from social media platforms. *Expert Systems with Applications* **2023**, *223*, 119862.

2. Hande, A.; Priyadarshini, R.; Chakravarthi, B.R. KanCMD: Kannada CodeMixed Dataset for Sentiment Analysis and Offensive Language Detection. In Proceedings of the 3rd PEOPLES Workshop. ACL, 2020, pp. 54–63.
3. Ngo, V.M.; Le-Khac, N.A.; Kechadi, M.T. Designing and Implementing Data Warehouse for Agricultural Big Data. In Proceedings of the Big Data – BigData 2019, 2019, pp. 1–17.
4. Cao, T.H.; Ngo, V.M.; Hong, D.T.; Quan, T.T. A Named-Entity-Based Multi-Vector Space Model for Semantic Document Clustering. In Proceedings of the Proceedings of PAKDD Workshop (WMWA'2008, Japan), 2008.
5. Duong, T.; Vu, T.; Ngo, V. Detecting Vietnamese Opinion Spam. In Proceedings of the Proceedings of Scientific Researches on the Information and Communication Technology in 2012 (ICTFIT'12), 2012.
6. Tran, T.N.T.; Nguyen, L.K.N.; Ngo, V.M. Machine Learning based English Sentiment Analysis. *Journal of Science and Technology, Vietnam Academy of Science and Technology* **2014**, *52*, 142–155.
7. Nguyen, L.H.; Pham, N.H.T.; Ngo, V.M. Opinion spam recognition method for online reviews using ontological features. *Journal of Science, Special Issue: Natural Science and Technology* **2014**, *61*, 44–59.
8. Nguyen, P.T.; Le, L.T.; Ngo, V.M.; Nguyen, P.M. Using Entity Relations for Opinion Mining of Vietnamese Comments. *Journal of Science and Technology, Vietnam Academy of Science and Technology* **2014**, *52*, 120–132.
9. Wankhade, M.; et al. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review* **2022**, *55*, 5731–5780.
10. Bordoloi, M.; Biswas, S.K. Sentiment analysis: A survey on design framework, applications and future scopes. *Artif Intell Rev* **2023**, pp. 1–56.
11. Ezhilarasan, M.; et al. Sentiment Analysis On Product Review: A Survey. In Proceedings of the Proceedings of the 2019 ICCPEIC, 2019, pp. 180–192.
12. Ngo, V.M.; et al. Investigation, Detection and Prevention of Online Child Sexual Abuse Materials: A Comprehensive Survey. In Proceedings of the 2022 RIVF International Conference on Computing and Communication Technologies, 2022, pp. 707–713.
13. Dao, P.Q.; Nguyen-Tat, T.B.; Roantree, M.; Ngo, V.M. Exploring Multimodal Sentiment Analysis Models: A Comprehensive Survey. In Proceedings of the MAPR conf., 2024, pp. 1–7.
14. Capuano, N.; et al. Sentiment analysis for customer relationship management: an incremental learning approach. *Applied Intelligence* **2020**, *51*, 3339–3352.
15. Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *Journal of Big Data* **2021**, *8*.
16. Chakravarthi, B.R. Hope speech detection in YouTube comments. *Soc Netw Anal Min* **2022**, *12*, 75.
17. Dake, D.K.; Gyimah, E. Using sentiment analysis to evaluate qualitative students' responses. *Educ Inf Technol (Dordr)* **2023**, *28*, 4629–4647.
18. Kaplan, H.; Weichselbraun, A.; Brasoveanu, A.M.P. Integrating Economic Theory, Domain Knowledge, and Social Knowledge into Hybrid Sentiment Models for Predicting Crude Oil Markets. *Cognit Comput* **2023**, pp. 1–17.
19. Umair, A.; Masciari, E. Sentiment Analysis Using Improved CT-BERT_CONVLayer Fusion Model for COVID-19 Vaccine Recommendation. *SN Computer Science* **2024**, *5*.
20. Abimbola, B.; et al. Sentiment analysis of Canadian maritime case law: a sentiment case law and deep learning approach. *International Journal of Information Technology* **2024**, *16*, 3401–3409.
21. Kastrati, M.; et al. Leveraging distant supervision and deep learning for twitter sentiment and emotion classification. *J. Intell. Inf. Syst.* **2024**, *62*, 1045–1070.
22. Colon-Ruiz, C.; Segura-Bedmar, I. Comparing deep learning architectures for sentiment analysis on drug reviews. *J Biomed Inform* **2020**, *110*, 103539.
23. Rakhmanov, O. A Comparative Study on Vectorization and Classification Techniques in Sentiment Analysis to Classify Student-Lecturer Comments. *Procedia Computer Science* **2020**, *178*, 194–204. 9th YSC-2020,.
24. Iddrisu, A.M.; et al. A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector. *IJIM Data Insights* **2023**, *3*.
25. Spinde, T.; et al. What do Twitter comments tell about news article bias? Assessing the impact of news article bias on its perception on Twitter. *Online Social Networks and Media* **2023**, *37*-38.
26. Meena, G.; et al. A hybrid DL approach for detecting sentiment polarities and KG representation on monkeypox tweets. *Decision Analytics Journal* **2023**, *7*.
27. Atandoh, P.; et al. Integrated deep learning paradigm for document-based sentiment analysis. *J. King Saud Univ. - Comput. Inform. Sci.* **2023**, *35*.

28. George, M.; Murugesan, R. Improving sentiment analysis of financial news headlines using hybrid Word2Vec-TFIDF feature extraction technique. *Procedia Computer Science* **2024**, *244*, 1–8. 6th Int. Conf. on AI in Computational Linguistics.
29. Tanoto, K.; et al. Investigation of challenges in aspect-based sentiment analysis enhanced using softmax function on twitter during the 2024 Indonesian presidential election. *Procedia Computer Science* **2024**, *245*, 989–997. 9th ICCSCI-2024.
30. Greeshma, M.; Simon, P. Bidirectional Gated Recurrent Unit with Glove Embedding and Attention Mechanism for Movie Review Classification. *Procedia Computer Science* **2024**, *233*, 528–536. 5th ICIDCA-2024.
31. Taneja, K.; Vashishtha, J.; Ratnoo, S. Transformer Based Unsupervised Learning Approach for Imbalanced Text Sentiment Analysis of E-Commerce Reviews. *Procedia Computer Science* **2024**, *235*, 2318–2331. ICMLDE-2023.
32. Zhai, G.; et al. Multi-attention fusion modeling for sentiment analysis of educational big data. *Big Data Mining and Analytics* **2020**, *3*, 311–319.
33. Mishev, K.; et al. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access* **2020**, *8*, 131662–131682.
34. Fazil, M.; et al. Attentional Multi-Channel Convolution With Bidirectional LSTM Cell Toward Hate Speech Prediction. *IEEE Access* **2023**, *11*, 16801–16811.
35. Zhong, Q.; et al. Knowledge Graph Augmented Network Towards Multiview Representation Learning for Aspect-Based Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering* **2023**, *35*, 10098–10111.
36. Xiang, C.; et al. Phrase-Aware Financial SA Based on Constituent Syntax. *IEEE/ACM Tran. on Audio, Speech, and Lan. Processing* **2024**, *32*, 1994–2005.
37. Duan, G.; Yan, S.; Zhang, M. A Hybrid Neural Network Model for Sentiment Analysis of Financial Texts Using Topic Extraction, Pre-Trained Model, and Enhanced Attention Mechanism Methods. *IEEE Access* **2024**, *12*, 98207–98224.
38. Guan, M.; Li, F.; Xue, Y. Enhanced Syntactic and Semantic Graph Convolutional Network With Contrastive Learning for Aspect-Based Sentiment Analysis. *IEEE Transactions on Computational Social Systems* **2024**, *11*, 859–870.
39. Sherin, A.; Jasmine Selvakumari Jeya, I.; Deepa, S.N. Enhanced Aquila Optimizer Combined Ensemble Bi-LSTM-GRU With Fuzzy Emotion Extractor for Tweet Sentiment Analysis and Classification. *IEEE Access* **2024**, *12*, 141932–141951.
40. Xiao, H.; Luo, L. An Automatic Sentiment Analysis Method for Short Texts Based on Transformer-BERT Hybrid Model. *IEEE Access* **2024**, *12*, 93305–93317.
41. Perti, A.; Sinha, A.; Vidyarthi, A. Cognitive Hybrid Deep Learning-based Multi-modal Sentiment Analysis for Online Product Reviews. *ACM Transactions on Asian and Low-Resource Language Information Processing* **2023**.
42. Dang, C.N.; et al. Sentiment Analysis for Vietnamese – Based Hybrid Deep Learning Models. In Proceedings of the 18th Hybrid Artificial Intelligent Systems (HAIS), 2023, pp. 293–303.
43. Alzaidi, S.A.; et al. A Text-Inception-Based NLP for Sentiment Analysis of Drug Experiences. *ACM Trans. Asian Low-Resource Lang. Infomation Process.* **2024**.

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.