JOURNAL of INTERDISCIPLINARY KNOWLEDGE

DOI: https://doi.org/10.37497/jik.v8iknowledge.1637

ARTICLE

A Comprehensive Review of Current Methods, Progress, and Challenges in Sentiment Analysis

Uma Revisão Abrangente dos Métodos Atuais, Avanços e Desafios na Análise de Sentimentos

Shradha Balasaheb Linge
Maharashtra Institute of Technology (MIT), Pune (India).
E-mail: shradha.linge@gmail.com

Vaishali Umesh Sathe
Maharashtra Institute of Technology (MIT), Pune (India).
E-mail: vaishalisathe138@gmail.com

Prerana Swapnil Patil
Maharashtra Institute of Technology (MIT), Pune (India).
E-mail: patil.prerana@gmail.com

Sharmila More
Maharashtra Institute of Technology (MIT), Pune (India).
E-mail: ssmore@mitacsc.ac.in

ABSTRACT | **Purpose:** This paper reviews the current state, progress, and challenges of **Sentiment Analysis** (**SA**), a computational approach to identifying and interpreting human opinions and emotions from text. It aims to evaluate the evolution of methods, highlight persistent limitations, and explore emerging directions toward developing generalizable, explainable, and ethically responsible sentiment analysis systems. **Design/Methodology/Approach:** The review integrates evidence from two key sources: (1) a *tertiary study* (Lighart et al., 2021) summarizing outcomes from 14 systematic literature reviews and mapping studies, and (2) a *domain-specific review* (Sweta, 2024) on sentiment analysis in education. Through thematic analysis, the study identifies methodological trends, core challenges, and innovations shaping the discipline, linking large-scale findings with domain-level applications. **Findings:** The analysis reveals a significant methodological transition—from lexicon-based and traditional machine learning models (SVM, Naive Bayes) to **deep learning** and **transformer-based architectures** (CNN, LSTM, BERT, GPT). Despite progress, key issues persist: domain and language dependency, lack of contextual sensitivity (sarcasm, irony), data imbalance, model opacity, and privacy concerns. New directions, such as **cross-domain and multilingual modeling**, **explainable AI** (**XAI**), **multimodal analysis**, and **real-time SA**, demonstrate promising solutions. **Research Limitations/Implications:** The review highlights the need for cross-lingual datasets, explainable architectures, and interdisciplinary collaborations to enhance generalizability, fairness, and ethical compliance. **Originality/Value:** By combining high-level synthesis with domain-specific insights, this work provides a comprehensive and forward-looking framework for building robust, transparent, and socially responsible sentiment analysis models.

Keywords | Opinion Mining**,** Viewpoint Extraction**,** Generalization, Deep Structured Learning, NLP Systematic Review, Cross-Domain, Explainable AI.

Received: 14 September 2025
Revised: 14 November 2025
Accepted: 5 December 2025
e-ISSN: 3086-0016

Associate editor: Altieres Silva – Alumni.In Publisher

How to cite this article: Linge, S. B., Sathe, V. U., Patil, P. S., & More , S. (2025). A Comprehensive Review of Current Methods, Progress, and Challenges in Sentiment Analysis. *Journal of Interdisciplinary Knowledge*, 8(knowledge), e01637. https://doi.org/10.37497/jik.v8iknowledge.1637

RESUMO | Objetivo: Revisar de forma sistemática o estado da arte, os avanços recentes e os principais desafios da Análise de Sentimentos (AS), uma abordagem computacional voltada à identificação e interpretação de opiniões e emoções humanas a partir de textos. O estudo busca avaliar a evolução metodológica da área, identificar limitações persistentes e explorar direções emergentes para o desenvolvimento de sistemas generalizáveis, explicáveis e eticamente responsáveis. **Método:** A pesquisa adota uma abordagem de revisão teórica integrativa, combinando evidências de duas fontes centrais: (i) um estudo terciário que sintetiza resultados de revisões sistemáticas e estudos de mapeamento sobre Análise de Sentimentos e (ii) uma revisão específica sobre aplicações da AS no domínio educacional. Por meio de análise temática, são identificadas tendências metodológicas, desafios recorrentes e inovações tecnológicas, articulando resultados em nível macro com aplicações setoriais. **Resultados:** Os resultados evidenciam uma transição metodológica significativa, que vai de abordagens baseadas em léxicos e modelos clássicos de aprendizado de máquina para arquiteturas de *deep learning* e modelos baseados em *transformers*, como BERT e GPT. Apesar dos avanços, persistem desafios relevantes, incluindo dependência de domínio e idioma, limitações de sensibilidade contextual (sarcasmo e ironia), desequilíbrio de dados, opacidade dos modelos e preocupações éticas relacionadas à privacidade. Tendências emergentes, como modelagem multilíngue, análise multimodal, explicabilidade por meio de XAI e processamento em tempo real, apresentam potencial para superar tais limitações. **Conclusão:** Conclui-se que o futuro da Análise de Sentimentos depende da integração entre inovação técnica, interpretabilidade e princípios éticos, de modo a viabilizar sistemas robustos, transparentes e socialmente responsáveis.

Palavras-chave | Análise de sentimentos; Mineração de opinião; Aprendizado profundo; Processamento de linguagem natural; Generalização; Inteligência artificial explicável; Revisão sistemática.

1 INTRODUCTION

Digital platforms generate an enormous corpus of audience-authored content, such as consumer reviews and social media commentary and social media commentary, forum discussions, etc.—that forms a rich pool of human sentiment and opinion. Opinion mining, is the measured way that specializes in extracting and measuring these subjective states from text data [1, 2]. It is broadly used for a significant, game-changing applications, from monitoring brand perception and measuring political mood to tailoring education and improving customer service.

Early SA work tended to produce models that worked superbly on the particular dataset that they were trained on but collapsed when used in new settings. This failure to generalize arises due to the intrinsic variability of human language: sentiment is articulated variously across cultures (multilingualism), industries (domain-dependency), and social media platforms (context-dependency), and frequently through sophisticated linguistic mechanisms like sarcasm and irony [3, 4].

The main challenge thus is to go beyond thin, fragile models and toward strong, generalizable SA systems. This paper meets this challenge by performing an in-depth review of the progress in the field. We integrate findings from two main sources:

1. A tertiary study (Lighart et al., 2021) [1] which systematically compiles outcomes from 14 Systematic Literature Reviews (SLRs) and Systematic Mapping Studies (SMSs) on SA, presenting a high-level, evidence-based snapshot of the field.
2. A domain-specific review (Sweta, 2024) [2] that explores the new trends and specific challenges of using SA in educational technology as a close case study of generalization in a complicated domain.

By combining the wide view of the tertiary research with the intensive, practical emphasis of the educational review, this paper seeks to:

- Record the evolutionary history and state-of-the-art of opinion mining methods.
- Scan and examine the essential issues that prevent generalization.
- Pinpoint emerging trends and creative solutions that are expanding the horizons of what is achievable.
- Offer a coherent survey of future directions of research to deliver genuinely solid and generalizable SA.

2 LITERATURE REVIEW AND BACKGROUND

2.1 Defining Sentiment Analysis and Its Tasks

SA is not a single problem but a collection of related ones. Liu (2012) characterizes it as computational opinion, sentiment, emotion, and attitude study of people [3]. Typical tasks are:

- Sentiment classification is the task of determining whether a text's overall polarity is positive, negative, or neutral. [1, 4].
- Subjectivity Classification: Separating facts from subjective opinions [1, 5].
- Aspect-Based SA (ABSA): Determination of specific entities (e.g., "battery life" of a phone) and the sentiment expressed towards each of them [1, 6].
- Emotion Detection: Detection of specific emotions (e.g., happiness, anger, sorrow) as opposed to general polarity [2, 7].
- Opinion Spam Detection: Detection of deceptive or spam reviews [1].

2.2 Methodological Development

The research area has undergone a remarkable development in methodologies:

1. Dictionary-based approaches: The initial approaches were based on sentiment lexicons (e.g., VADER) that had words with their semantic orientation and intensity annotated. They look for these words in the text and score them based on rules (e.g., dealing with negations) [1, 2]. Their merit lies in domain adaptability without training data but their demerit is a failure to understand context.
2. Classic Machine Learning (ML) Methods: Early sentiment analysis was dominated by supervised models like SVM, Naive Bayes, and Logistic Regression, which relied heavily on textual features including BoW, n-grams, and TF-IDF for classification. for classification [1, 8]. They need extensive labeled corpora but tend to outperform early lexicon approaches.
3. Deep Learning (DL) Techniques: DL has transformed SA with the introduction of models such as CNNs, which are particularly good at extracting local features, whereas RNNs and their extensions (LSTM, GRU) are optimized to process sequence data as well as long-distance dependencies [1, 9].

- At the forefront of research are transformer-based architectures such as BERT. Transformer-based Models (BERT, GPT): Trained on massive corpora, these models produce context-sensitive word embeddings that significantly enhance performance across tasks through understanding word meaning in context [2, 10].
- Attention Mechanisms: Enable models to selectively pay attention to the most salient components of the input text for making a decision, important in cases such as ABSA [1, 11].

Table 1. Evolution of Key SA Approaches

Era	Primary Approach	Key Techniques	Strengths	Weaknesses
Early	Lexicon-Based	VADER, SentiWordNet	No training data needed, interpretable	Poor with context, sarcasm, domain-specific slang
2000s-2010s	Traditional ML	SVM, Naive Bayes, NB	Effective with good features, better than lexicons	Manual feature engineering, struggles with complexity
2010s-Present	Deep Learning	CNN, RNN, LSTM, GRU	Automatic feature extraction, high accuracy	Data-hungry, computationally expensive, "black box"
Present-Future	Pre-trained Transformers	BERT, GPT, XLNet	State-of-the-art, contextually aware, generalizable	Very resource-intensive, complex fine-tuning needed

2.3 The Need for a Tertiary Study

With the growth of SA research, secondary studies (SLRs and SMSs) now exist to synthesize the knowledge. The tertiary study synthesizes these secondary studies to offer an even greater level of synthesis, capturing overarching patterns, consensus, and gaps for the whole field [1]. The Lighart et al. (2021) [1] tertiary study is the core evidence base for this review, making our analysis reliant on aggregated findings from hundreds of primary studies.

3 METHODOLOGY

The review uses a systematic method for synthesizing the knowledge of the two source documents provided:

1. Source Analysis: The heart of the present review is a synthesis of Lighart et al.'s tertiary study [1] and Sweta's chapter on educational SA [2]. The former gives a bird's eye view of the field's characteristics, methodologies, datasets, and problems. The latter gives a close-up view of how these generic problems occur and are tackled in one particular, challenging field (education).
2. Thematic Analysis: The repeated themes from both reports were coded, categorized, and termed as major categories: Techniques, Challenges, and Emerging Directions. This enabled cross-cutting analysis that bridges the broad field-wide trends with precise domain applications.

3. Integration: Results from both sources were combined to create a coherent account. For instance, the "domain dependency" issue revealed in the tertiary work [1] is made tangible with real-life examples from the review of the education domain [2].
4. Advances in Techniques and Models

There has been substantial advancement in the methods employed for SA, in significant part thanks to the transition to deep learning.

4.1 The Supremacy of Deep Learning

The tertiary research verifies that DL, and most notably LSTMs and CNNs, are the most popular algorithms in recent SA research [1]. Examination of 112 articles in 2020 identified LSTM (35.53%) and CNN (33.33%) as the leading two architectures, followed by GRU (8.77%) and RNN (7.89%) [1, Table 3]. Transformer models such as BERT, albeit not prevalent within that particular 2020 time slice, are acknowledged as being current state-of-the-art due to their pre-trained contextual knowledge [2].

4.2 Hybrid and Ensemble Methods

To tap the best of various models, hybrid methods are gaining popularity. These are:

- Lexicon-ML Hybrids: Integrating linguistic rules of lexicons with pattern detection of ML models to enhance feature sets [1].
- DL Architecture Hybrids: Many effective and popular hybridization patterns involve stacking CNNs (for feature extraction) and LSTMs (for sequence modeling) [1, 2]. For instance, CNN-LSTM models are widely used in educational SA for the analysis of student feedback [2].
- Symbolic-Subsymbolic Hybrids: Coupling symbolic AI (knowledge graphs, ontologies such as SenticNet) with subsymbolic DL models to leverage commonsense knowledge and enhance interpretability [1, 12].

4.3 Application in Specific Domains: The Case of Education

The education sector is a perfect example of the application of these cutting-edge methods. Sweta (2024) illustrates the application of SA to:

- Learning System Evaluation: Evaluating feedback on MOOCs and online courses with ensemble learning and LSTMs [2].
- Improving Pedagogy: Topic modeling with LDA and reinforcement learning to adjust educational material according to student sentiment [2].
- Evaluation of Assessments: Utilizing methods such as VADER and Random Forests to measure the sentiment of students towards exams and assignments [2].

This illustrates how general-purpose SA models are being adapted and applied effectively to address domain-specific issues.

5 KEY CHALLENGES IN GENERALIZING SENTIMENT ANALYSIS

While there has been astounding progress, various challenges hinder SA models from generalizing truly.

5.1 Domain and Language Dependency

This is the most commonly mentioned difficulty in the tertiary study [1]. Product-review-trained models (e.g., Amazon) do poorly on social media content (e.g., Twitter) because of vocabulary, style, and syntactic differences. Correspondingly, an English-trained model does poorly on Arabic or Chinese unless adapted. This creates expensive and time-consuming label dataset creation for each new language or domain [1, 2, 13].

5.2 Contextual Sensitivity and Subtlety

Human emotion is contextually sensitive. Some of the challenges are:

- Sarcasm and Irony: An otherwise positive term with a negative context (e.g., “Oh, great! Another Monday.”) Inverts the sentiment, sometimes tricking models [1, 2].
- Comparative Sentences: Recognizing that “Product A is better than Product B” means A is positively framed and B is comparatively negatively framed [4].
- Domain-Specific Meaning: The term “sick” is bad in the context of medicine but good in slang review contexts (“That movie was sick!”) [1].

5.3 Data Imbalance and Bias

Reviews in datasets tend to consist much more often of positive reviews than negative ones, causing the models to be biased towards the majority class. In addition, models may inherit and exacerbate societal biases in the training data, resulting in discriminatory or unfair outcomes [2, 14].

5.4 Model Interpretability and Explainability

Deep learning models tend to be “black boxes,” so that it is hard to know why a certain sentiment classification was produced. The lack of explainability in these models severely hinders their trustworthiness and adoption, especially in critical fields like healthcare or finance [2, 15]. Explainable AI (XAI) is becoming a key sub-field to help overcome this.

5.5 Data Privacy and Ethics

SA frequently handles personal and sensitive textual information. Protection of user privacy by anonymization of data, following regulations such as GDPR, and formulating ethical standards to use SA outcomes are key concerns [2, 16]. This is particularly important in educational institutions where data concerns children [2].

5.6 Scalability and Computational Efficiency

Processing massive datasets (e.g., millions of tweets) is computationally expensive and time-intensive. Efficient algorithms and the use of distributed computing frameworks are required for practical applications [2].

6 EMERGING TRENDS AND DIRECTIONS

The area is actively developing to meet these challenges. Sweta (2024) and Lighart et al. (2021) both present multiple promising new directions [1, 2].

6.1 Cross-Domain and Multilingual SA

Research centers on methods that enable knowledge transfer from a data-abundant "source domain" (or language) to a data-poor "target domain." Techniques involve algorithms for domain adaptation and leveraging pre-trained multilingual models such as mBERT that support multiple languages out-of-the-box [1, 2, 17].

6.2 Explainable AI (XAI) for SA

There has been increasing demand for models that do not just excel in terms of performance but can also justify their decisions. This includes creating visualizations or natural language descriptions indicating which words or phrases had the greatest impact on the sentiment prediction, building trust and allowing for debugging [2, 15].

6.3 Multimodal Sentiment Analysis

Text is usually supported by other modalities. Multimodal SA combines text with visual information (images, video frames) and sound (tone of voice) to reach a more comprehensive and precise interpretation of sentiment. For instance, a sarcastic tweet can be supported by an eye-rolling emoji or a sarcastic photo [2, 18].

6.4 Real-Time Sentiment Analysis

The capacity to process sentiment analysis in real-time data streams (for instance, in a live lecture, product release, or political forum) enables dynamic response and instant feedback. This necessitates highly scalable and efficient models [2].

6.5 Deep Learning Architectures

Application of more advanced models is increasingly prevalent:

- Transformers and BERT: Pre-trained transformer model fine-tuning for particular SA tasks is still the most favorable method for achieving high precision [2, 10].
- Graph Neural Networks (GNNs): Employed to represent word or aspect dependencies in a sentence, retaining syntactic and semantic interdependencies that shallow models neglect [1].
- Attention Mechanisms: More sophisticated (e.g., self-attention, hierarchical attention) to effectively capture intricate text structures [1, 11].

6.6 Educational and Learning Analytics Integration

In the educational field, SA is never applied separately. Instead, the pattern is to combine it with other data sources (such as grades, clickstream data, social network analysis) in learning analytics systems to understand the totality of the learner experience and deliver customized interventions [2, 19].

7 DISCUSSION

The path towards generalizable SA is a progression from inflexible, context-insensitive models to dynamic, context-sensitive systems. The tertiary study [1] gives the empirical foundation, demonstrating the community's agreement about the predominance of DL and the ubiquity of issues such as domain dependency. The education-oriented review [2] adds a human face to these issues, demonstrating how they have direct implications for the efficacy of technology-enhanced learning.

1. The important realization is that there is no one "silver bullet." Generalization needs to happen through a multi-pronged strategy:
2. Architectural Innovation: Relying on pre-trained, context-sensitive models (Transformers) as a solid anchor.
3. Methodological Innovation: Creating improved methods of transfer learning, domain adaptation, and explainability.
4. Ethical Forethought: Infusing privacy and fairness into the SA system design from scratch.

5. Domain Collaboration: Tightly integrated collaboration between NLP researchers and domain specialists (e.g., educators, psychologists) to comprehend the subtleties of sentiment in certain contexts.

The trends are such that SA systems will become more accurate, transparent, multimodal, and ethically sound in the future, able to operate reliably over the enormous and diverse universe of human expression.

8 CONCLUSION AND FUTURE WORK

This review has taken a current snapshot of sentiment analysis, drawing on extensive tertiary research and in-depth domain-specific examination to set out the direction toward generalization. We have charted the obvious advancement from lexicon-based techniques to dominant deep learning structures that now constitute the state of the art. There are still important challenges, though, chiefly concerned with context, domain, bias, and explainability.

The future directions look promising. The future of generalized SA is in:

- Stronger Foundation Models: Larger, better-performing, and more multilingual pre-trained models.
- Causal and Explainable Models: Beyond correlation to seeing the causal processes driving sentiment expression and model reasoning made explicit.
- Human-in-the-Loop Systems: Building systems where AI does large-scale analysis and humans offer guidance and do complex edge cases.
- Standardized Ethical Frameworks: Establishing and embracing industry standards for the ethical collection, annotation, and use of sentiment data.

By directing efforts towards these fields, the research community will be able to overcome existing limitations and develop SA systems that are not only technologically advanced but also resilient, equitable, and genuinely useful in all walks of human life.

9 REFERENCES

[1] Cambria, E., et al.: SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 105–114 (2020).

[2] Dashtipour, K., et al.: Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. Cognitive Computation. 8(4), 757–771 (2016). https://doi.org/10.1007/s12559-016-9415-7

[3] Devlin, J., et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186 (2019).

[4] Evans, D.: Emotion: The science of sentiment. American Journal of Orthopsychiatry. 72(4), 133–144 (2002).

[5] Fiok, K., Farahani, F.V., Karwowski, W., Ahram, T.: Explainable artificial intelligence for education and training. The Journal of Defense Modeling and Simulation. 19(2), 133–144 (2022). https://doi.org/10.1177/1548512920981595

[6] Hajrizi, R., Nucj, K.P.: Aspect-Based Sentiment Analysis in Education Domain. arXiv preprint arXiv:2010.01429 (2020).

[7]Kaur, G., Gujrati, R., & Uygun, H. (2023). How does AI fit into the management of human resources?. *Review of Artificial Intelligence in Education*, *4*(00), e4. https://doi.org/10.37497/rev.artif.intell.education.v4i00.4

[7] Kasmuri, E., Basiron, H.: Subjectivity analysis in opinion mining—A systematic literature review. International Journal of Advanced Soft Computing Applications. 9(3) (2017).

[8] Liang, W., et al.: Advances, challenges and opportunities in creating data for trustworthy AI. Nature Machine Intelligence. 4(8), 669–677 (2022). https://doi.org/10.1038/s42256-022-00516-1

[9] Lighart, A., Catal, C., Tekinerdogan, B.: Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review. 54, 4997–5053 (2021). https://doi.org/10.1007/s10462-021-09973-3

[10] Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. 5(1), 1–167 (2012).

[11] Poria, S., et al.: Multimodal Sentiment Analysis: Addressing Key Issues and Setting up the Baselines. IEEE Intelligent Systems. 33(6), 17–25 (2018). https://doi.org/10.1109/MIS.2018.2880005

[12]Prokhorova, Y., Gujrati, R., & Uygun, H. (2024). The use of AI Chatbots in higher education: the problem of plagiarism. *Review of Artificial Intelligence in Education*, *5*(00), e031. https://doi.org/10.37497/rev.artif.intell.educ.v5i00.31

[12] Qazi, A., et al.: A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges. Internet Research. 27(3), 608–630 (2017). https://doi.org/10.1108/IntR-04-2016-0086

[13] Romero, C., Ventura, S.: Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery. 10(3), e1355 (2020). https://doi.org/10.1002/widm.1355

[14] Silbar, B., Zlatic, L., R., N.: Ethical and Privacy Issues of Learning Analytics in Higher Education. In: INTED2021 Proceedings, pp. 7766–7775 (2021).

[15] Sweta, S.: Emerging Trends and Challenges in Educational Sentiment Analysis. In: SpringerBriefs in Applied Sciences and Technology. Springer, Singapore (2024). https://doi.org/10.1007/978-981-87424-1_5

[16] Wang, Y., et al.: Attention-based LSTM for Aspect-level Sentiment Classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615 (2016).

[17] Wankhade, M., Rao, A.C.S., Kulkarni, C.: A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review. 55(7), 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10138-z

[18] Zhang, W., et al.: A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. IEEE Transactions on Knowledge and Data Engineering. 35(11), 11019–11038 (2023). https://doi.org/10.1109/TKDE.2022.3232985

[19] Zhang, Y., Liu, B., Wang, S.: Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8(4), e1253 (2018). https://doi.org/10.1002/widm.1253