Natural Language Processing Journal 6 (2024) 100059

Contents lists available at ScienceDirect

Natural Language Processing Journal

journal homepage: www.elsevier.com/locate/nlp

Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review

Jamin Rahman Jim [a], Md Apon Riaz Talukder [b], Partha Malakar [b], Md Mohsin Kabir [c], Kamruddin Nur [b], M.F. Mridha [b,*]

[a] Advanced Machine Intelligence Research Lab, Banani, Dhaka, 1213, Dhaka, Bangladesh
[b] Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh
[c] Faculty of Informatics, Eötvös Loránd University, H-1117 Budapest, Hungary

ARTICLE INFO

Keywords:
Sentiment classification
Text classification
Natural language processing
Emotion detection
Sentiment analysis

ABSTRACT

Sentiment analysis is a method within natural language processing that evaluates and identifies the emotional tone or mood conveyed in textual data. Scrutinizing words and phrases categorizes them into positive, negative, or neutral sentiments. The significance of sentiment analysis lies in its capacity to derive valuable insights from extensive textual data, empowering businesses to grasp customer sentiments, make informed choices, and enhance their offerings. For the further advancement of sentiment analysis, gaining a deep understanding of its algorithms, applications, current performance, and challenges is imperative. Therefore, in this extensive survey, we began exploring the vast array of application domains for sentiment analysis, scrutinizing them within the context of existing research. We then delved into prevalent pre-processing techniques, datasets, and evaluation metrics to enhance comprehension. We also explored Machine Learning, Deep Learning, Large Language Models and Pre-trained models in sentiment analysis, providing insights into their advantages and drawbacks. Subsequently, we precisely reviewed the experimental results and limitations of recent state-of-the-art articles. Finally, we discussed the diverse challenges encountered in sentiment analysis and proposed future research directions to mitigate these concerns. This extensive review provides a complete understanding of sentiment analysis, covering its models, application domains, results analysis, challenges, and research directions.

Contents

* Corresponding author.
E-mail address: firoz.mridha@aiub.edu (M.F. Mridha).

https://doi.org/10.1016/j.nlp.2024.100059
Received 14 November 2023; Received in revised form 31 January 2024; Accepted 18 February 2024

1. Introduction

Sentiment analysis, frequently called opinion mining, constitutes a pivotal component of Natural Language Processing (NLP) (Liu, 2022). Its primary function is to systematically identify and categorize sentiments embedded within text data, enabling the assessment of emotional tones, opinions, and attitudes conveyed through written or spoken language. Sentiment analysis techniques encompass a range of approaches, beginning with the primary classification of text into positive, negative, or neutral sentiments. However, its capabilities extend to more intricate methods, permitting the discernment of specific emotions, intentions, or nuanced aspects of sentiment, such as joy, anger, sarcasm, or context-specific sentiments observed in domains like product reviews (Cambria et al., 2017). The general process of a sentiment analysis system encompasses stages such as data collection, text preprocessing, feature extraction, training machine learning or deep learning models, and thorough evaluation to assess the model's effectiveness in sentiment classification tasks (Devika et al., 2016). Here, Fig. 1 illustrates the working process of a sentiment analysis model. These collective procedures empower organizations and researchers to extract valuable insights from textual content, facilitating informed decision-making and personalized responses to the sentiments expressed in text data.

Sentiment analysis, a crucial aspect of natural language processing, holds significant value and presents numerous advantages. It empowers organizations to glean valuable insights from public opinions and customer feedback, facilitating data-driven decision-making, product enhancement, and effective marketing strategies (Ahmed et al., 2022). By automatically categorizing sentiments into positive, negative, or neutral, sentiment analysis simplifies the analysis of extensive textual data, proving indispensable for businesses aiming to comprehend customer sentiment, manage online reputation, and stay abreast of market trends. Furthermore, sentiment analysis shapes social and political discussions, aiding researchers and policymakers in understanding public sentiment on vital issues and ultimately fostering more informed and responsive decision-making in an increasingly digital and interconnected world (Peng et al., 2022).

Sentiment analysis has experienced notable progress in recent years, primarily propelled by utilizing machine learning (ML) (Revathy et al.,

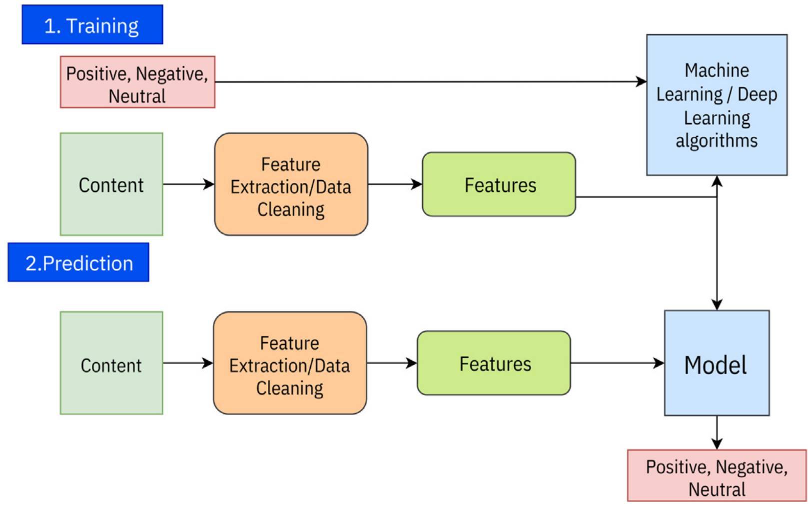

Fig. 1. This figure illustrates the working process of sentiment analysis.

2022) and deep learning (DL) (Abdullah and Ahmet, 2022) techniques in sentiment classification. These techniques, encompassing both traditional ML algorithms and advanced deep neural networks, have significantly improved the accuracy and scalability of sentiment analysis systems, leading to the creation of more sophisticated models. Thus, thoroughly examining the current state-of-the-art developments in sentiment analysis, covering application domains, datasets, pre-processing techniques, ML, DL, LLMs and pre-trained models, and analyzing recent experimental results with their limitations, is imperative for a comprehensive understanding of the field's recent advancements (Hussein, 2018). While existing surveys exist, none comprehensively encompass all these aspects, indicating a notable research gap.

Our Motivation and Objective: To provide a better understanding of the current state-of-the-art advancement of sentiment analysis we conducted this review article by specifically focusing on the recent research articles, their application domain, and experimental analysis in sentiment analysis. Briefly, in this article, we dive into diverse applications of sentiment analysis, commonly employed datasets, preprocessing methodologies, and ML, DL, LLMs, and pre-trained models. Additionally, the survey analyzed the experimental findings of recently published research articles, identified existing challenges, and proposed future research directions to address these challenges.

The novelty of our article compared to existing articles: Our review article is more informative and novel than existing surveys in several ways. Most existing surveys discuss the general working process and types of sentiment analysis. In contrast, our survey focuses on the applications of ML, DL, LLMs, and Pre-trained-based sentiment analysis in diverse areas by analyzing existing articles. Additionally, our survey discusses the commonly used datasets, preprocessing methods and ML, DL, LLMs, and Pre-trained models used in sentiment analysis. While some existing survey only discusses recent articles, our survey presents the experimental results of state-of-the-art articles, and also by analyzing their limitations, our survey offers a better understanding of the advancements of sentiment analysis. Furthermore, Table 1 compared existing surveys with ours. This article will assist investors and researchers in developing the ML, DL, LLMs, and Pre-trained models based sentiment analysis ecosystem and understanding its progress.

Briefly, this paper's significant contributions are summarized below.

- An in-depth investigation into the various application domains of sentiment analysis using ML, DL, LLMs and Pre-trained models. This exploration delves into researchers' progress in this field, encompassing a broad spectrum of practical areas where sentiment analysis finds its applications.
- Provides an overview of commonly used sentiment analysis datasets, data pre-processing techniques and evaluation metrics for optimal model training.
- A thorough exploration of the key ML, DL, LLMs and Pre-trained models employed in sentiment analysis, including discussion on multimodal sentiment analysis and labels perspective in sentimental analysis.
- A comprehensive examination that scrutinizes recent advancements and contributions by researchers, offering insights into the state-of-the-art experimental results that influence the landscape of the subject.
- Engaging in a comprehensive discussion covering an extensive range of ML, DL, LLMs and Pre-trained-based sentiment analysis challenges, outlining future research opportunities to confront and alleviate these impediments.

The paper's structure is as follows: We described the methodology in Section 2, explored various application domains of sentiment analysis in Section 3, and then delved into commonly used datasets and preprocessing methods in sentiment analysis in Sections 5.1 and 5.2. Sections 4.1 and 4.2 discuss the multimodal sentiment analysis and labels perspective in sentimental analysis, respectively. In Section 6, we briefly described ML, DL, LLMs and pre-trained models utilized in sentiment analysis. Next, in Section 7, we analyzed recent state-of-the-art experimental results, and Section 8 outlined significant sentiment analysis challenges and research opportunities for improvement. Finally, we concluded the paper in Section 9.

2. Methodology

The survey employs the Systematic Literature Review (SLR) approach based on the methodology introduced by Keele et al. (2007), Kitchenham (2004). This paper delineates the SLR procedure across three clearly defined phases: identifying research questions, defining criteria for inclusion and exclusion for review, and finally selecting the articles for the review.

2.1. Research questions

The primary research questions were:
RQ1: What is the working procedure of sentiment analysis?
RQ2: What are the areas of application for sentiment analysis?

Table 1
Comparative analysis of existing review articles with our article.

Ref.	Application	Datasets	Pre-Processing	Algorithm	Results analysis	Challenges & Future work	Novelty
Wankhade et al. (2022)	✓	✗	✓	✓	Analyzed article published before 2020	✓	Provided an overview of sentiment analysis methods, applications, and challenges, along with a comparative evaluation of these methods, to aid in understanding their strengths, limitations and future research directions in this field.
Ligthart et al. (2021)	✓	✗	✗	✓	–	✓	Conducted a study in the field of sentiment analysis by synthesizing the results of previously published secondary studies, including systematic literature reviews and mapping studies.
Xu et al. (2022)	Mainly Social Media based	✓	✗	✓	Only Social Media Based	✓	Provided a systematic survey of social media-based sentiment analysis, highlighting new trends and challenges in the field.
Lu et al. (2023)	✓	✗	✗	✓	✗	✓	Covered various modalities and offered insights into recent advances in single-modal sentiment analysis and explored advanced studies on multimodal alignment and ChatGPT in SA, and discussed open research challenges and avenues for improvement.
Ma et al. (2023)	✓	✗	✗	✗	✗	✓	Provided an overview of the multi-modal sentiment analysis, with a specific focus on the transition from narrative sentiment to interactive sentiment, addressing the scarcity of literature in the realm of sentiment interaction.
This paper	✓	✓	✓	✓	✓	✓	Provides a comprehensive analysis and systematic review covering all the mentioned areas in ML, DL, LLMs and Pre-trained-based sentiment analysis by focusing on the most recent state-of-the-art research articles.

RQ3: Which research papers on sentiment analysis were published from 2020 to 2023,

and what application domains do they cover?

RQ4: Which datasets are commonly used for sentiment analysis experiments?

RQ5: What pre-processing techniques are typically employed in sentiment analysis?

RQ6: What are the commonly utilized ML, DL, LLMs, and Pre-trained models used in sentiment analysis,

and what are the computational techniques in sentiment analysis?

RQ7: What are the state-of-the-art advancements and experiments done in the field of sentiment analysis

from 2020 to 2023, and what are their limitations?

RQ8: What are the challenges and future research opportunities on sentiment analysis?

2.2. Criteria for inclusion and exclusion

The survey's essential resources were acquired by adhering to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Page et al., 2021), depicted in Fig. 2. Furthermore, the inclusion and exclusion criteria specified in PRISMA are in Table 2, outlining the standards used to evaluate papers for inclusion or exclusion.

At the beginning of this research, we selected 544 papers for the review. Table 3 shows the different keywords that we used for article selection in different databases and the number of articles we selected for the review for each keyword. These articles were selected in 4 different periods. They are September 2022 to October 2022, February 2023 to March 2023, 4th August 2023 to 12th September 2023, and a final selection was done from 2nd January 2024 to 11th January 2024.

2.3. Source of review articles

This survey focuses on academic articles of high quality retrieved from reputable databases, including ScienceDirect, SpringerLink, ACM Digital Library, IEEE Xplore, and well-recognized conferences. After a thorough evaluation, 162 papers focusing on ML, DL, LLMs and Pre-trained-based sentiment analysis were ultimately selected for review. Fig. 3 visualizes the bibliometric analysis of our selected papers based on their scopus index. Our deliberate choice to consider recent articles exclusively underscores our commitment to providing an advanced and state-of-the-art review.

3. Application

Sentiment analysis has diverse applications across industries. It offers valuable insights into public sentiment and opinions. Critical applications include customer feedback analysis for improving products and services, real-time social media monitoring for brands and individuals, market research to assess consumer trends, brand reputation management to prevent crises, political campaigning for understanding public opinion, and aiding product development. It is also used in financial markets, healthcare, and media outlets to gauge content impact and academic research to study public attitudes and social trends. Sentiment analysis is vital in understanding and responding to public assessing sentiment in the contemporary data-driven landscape across diverse domains. Fig. 4 illustrates the diverse applications of sentiment analysis.

3.1. Healthcare

Sentiment analysis is valuable in healthcare for monitoring patient feedback, online reviews, and public sentiment (Singh and Singh,

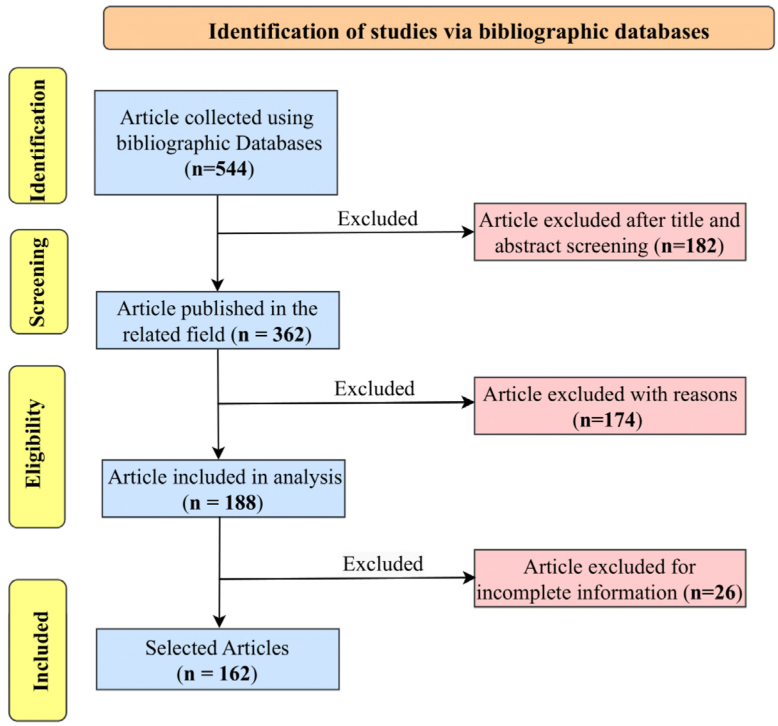

Fig. 2. PRISMA diagram depicting the article selection process for applications and the latest advancements in the field.

Table 2
The table discusses the including and excluding criteria for selecting articles.

	Inclusion Criteria	Exclusion criteria
Types of study	Original and review articles.	Thesis, white papers, communication letters, reports and editorials.
Language	Research articles written in English.	Duplicate and non-English articles.
Publication year	Articles published in 2020–2024 (For applications and Results analysis part).	Not related to the theme of the review.
Source	Articles published only in academic journals and conferences.	Articles that lack information and review papers.
Intervention	ML, DL, LLM and Pre-trained methods.	Traditional and statistical methods.
Region	Not restricted to a particular region.	–
Settings	Sentiment analysis.	Not related to sentiment analysis.

Table 3
The table discusses the keywords that we used for article selection in different databases and the paper count for respective keywords.

Keyword	Paper count
Machine learning based Sentiment+Emotion analysis	88
Deep learning based Sentiment+Emotion analysis	115
LLMs based Sentiment+Emotion Analysis	76
Sentiment analysis in healthcare	27
Sentiment analysis in financial markets	26
Sentiment analysis in customer support	34
Sentiment analysis in entertainment	25
Sentiment analysis in education	21
Sentiment analysis in E-commerce	23
Sentiment analysis in social media	26
Sentiment analysis in product reviews and consumer feedback	15
Multimodal sentiment analysis	43
Labels perspective in sentimental analysis	25

2023). It helps healthcare providers enhance patient satisfaction, make data-driven decisions, and understand public perception of healthcare policies and innovations. This technology helps improve patient care and inform health policy decisions, making it a critical tool in healthcare. Table 4 spotlights recent use cases of sentiment analysis in the healthcare sector.

Table 4
Summary of applications of sentiment analysis in healthcare.

Ref.	Description
Al-Mashhadany et al. (2022)	Using sentiment analysis of Facebook comments, classified Iraqi beauty centers as healthy or unhealthy.
Bansal and Kumar (2022)	Conducted an aspect-based analysis to compare hospitals on four key dimensions ("Physician Services", "Staff Services", "Hospital Facilities", and "Accessibility"), resulting in a comprehensive and descriptive hospital rating system.
Ghosh et al. (2023)	Created a system for identifying pain sensations by analyzing facial expressions. They used a tree-structured sub-model to extract characteristic patterns from the facial region through statistical and deep learning-based feature analysis.
Saranya et al. (2020)	Applying SVM-based sentiment analysis to healthcare-related tweets originating from Twitter, enabling real-time disease impact prediction and identifying key geographic areas discussing specific diseases. It used the Twitter API and Twitter4j for data extraction and preprocessing, enhancing healthcare analysis through social parameters.
Saad et al. (2021)	The paper applied and evaluated a hybrid sentiment analysis method for drug reviews, combining dictionary-based approaches (AFFIN, TextBlob, VADER) with machine learning models (LR, AB, RF, ETC, MLP) and various feature engineering techniques (TF, TF-IDF, TF U TF-IDF) in the medical domain.
Okey et al. (2023)	LLMs, like GPT and roBERTa, underpin advanced capabilities enabling nuanced opinion analysis and sentiment extraction from user text.
Thirunavukarasu et al. (2023)	Used LLMs in medicine, addressing challenges like errors and fabricated facts. It advocates for uncertainty indicators, clinician involvement, and accuracy improvement to enhance their reliability.
Tarcar et al. (2019)	Leveraging pretraining in NLP for enhanced Electronic Health Record extraction, the paper introduced a novel approach combining transfer learning and annotation tools, contributing to improved Named Entity Recognition model performance.

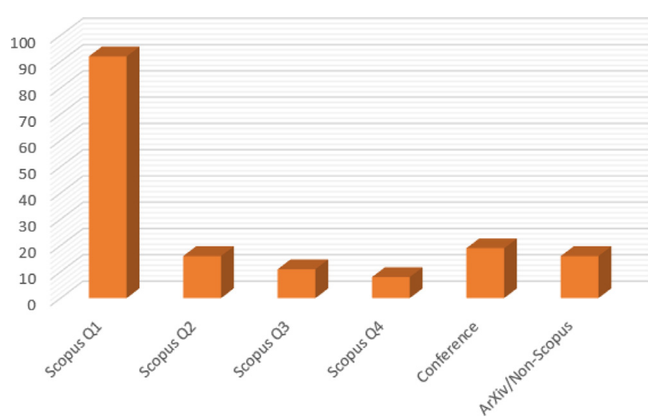

Fig. 3. Number of articles based on Scopus index.

3.2. Financial markets

Sentiment analysis is integral to financial markets, allowing investors and institutions to extract valuable information from the sentiment expressed in textual data. It helps in informed decision-making, risk assessment, and asset management by monitoring and interpreting emotions in sources such as news, social media, and reports (Mishev et al., 2020a). Table 5 presents some of the recent applications of sentiment analysis in the financial market.

3.3. Customer support

In customer support, sentiment analysis is vital for enterprises to oversee and regulate customer sentiment and satisfaction. By analyzing customer interactions such as support tickets, chat logs, and social media conversations, sentiment analysis helps organizations identify customer problems, assess service quality, and respond to customer feedback immediately (Diekson et al., 2023). This application helps improve customer experiences, reduce churn, and enhance customer relationship management. Diverse applications of sentiment analysis in customer support are discussed in Table 6.

3.4. Entertainment

Sentiment analysis is essential in the entertainment industry, allowing content creators, studios, and distributors to gauge audience reactions and preferences. By analyzing sentiment in user reviews, social media discussions, and audience comments, entertainment professionals can assess the reception of movies, TV shows, music, and other forms of entertainment. This application helps tailor content, improve marketing strategies, and predict box office success or viewership rates, ultimately helping to create more engaging and profitable entertainment experiences. Some recent applications of sentiment analysis in entertainment are discussed in Table 7.

3.5. Education

In education, the analysis of content and user-generated feedback provides valuable information on the effectiveness of teaching methods, learning materials and educational platforms. By examining the sentiments expressed in student evaluations, online discussions and educational forums, educators and institutions can assess student satisfaction, identify areas for improvement and adapt educational content to meet the diverse needs of students (Zhou and Ye, 2023). This application helps improve education quality, promote student engagement and optimize the learning experience. Table 8 demonstrates the application of sentiment analysis in the education domain.

3.6. E-commerce

In e-commerce, sentiment analysis is applied by analyzing product reviews to assess customer satisfaction, allowing for better product creation and marketing (Huang et al., 2023). In addition, it is used in social media monitoring to track brand sentiment, helping businesses adjust marketing strategies and respond more effectively to customer feedback, ultimately improving the overall shopping experience. Table 9 highlights some of the recent applications of sentiment analysis in e-commerce.

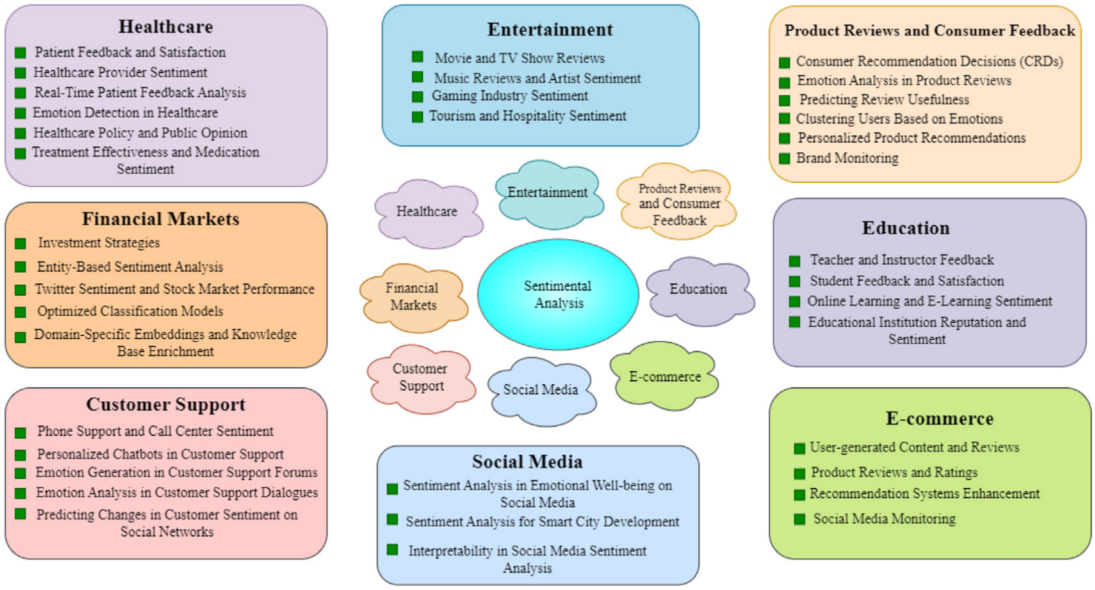

Fig. 4. This figure describes the diverse applications of sentiment analysis.

Table 5
Summary of applications of sentiment analysis in financial markets.

Ref.	Description
Guo et al. (2022)	Introduced a novel investment strategy leveraging deep reinforcement learning that integrates sentiment analysis with knowledge graphs.
Ahangari and Sebti (2023)	Presented an innovative hybrid method for sentiment analysis in social networks related to the Persian stock market using a dictionary-based approach and SVM evaluation.
Valle-Cruz et al. (2022)	Valle-Cruz investigated the connection between Twitter sentiment and stock market performance during the COVID-19 pandemic, employing semantic computing techniques.
Yekrangi and Nikolov (2023)	Multiple classification models were used in sentiment analysis for financial markets, demonstrating improved performance through optimized integration levels and heterogeneous text sources.
Mishev et al. (2020b)	Designed and implemented an evaluation platform, conducting over a hundred experiments on financial datasets that experts had identified. Demonstrated the increased effectiveness of contextual embeddings over traditional methods and the practicality of distilled NLP transformers for high-precision sentiment analysis in finance.
Yekrangi and Abdolvand (2021)	Developed a hybrid dictionary adapted for sentiment analysis in financial markets. It emphasized the need for continuous developments in datasets and sentiment analysis algorithms to capture market dynamics better.
Fatouros et al. (2023)	ChatGPT's implementation in financial sentiment analysis surpasses established models, showcasing the transformative impact of large language models in enhancing sentiment analysis within financial applications.
Araci (2019)	FinBERT, a BERT-based language model, leverages pre-training and domain-specific fine-tuning to excel in financial sentiment analysis, overcoming challenges posed by specialized language and limited labeled data in the financial domain.

3.7. Social media

Analyzing sentiments in social media entails applying natural language processing and machine learning methods to examine text-based data, including posts, comments, and messages on social platforms. The goal is to discern the emotion or emotional tone conveyed by users. Through the assessment of language and context, sentiment analysis categorizes content into positive, negative, or neutral sentiments, offering valuable insights into audience opinions and emotions regarding diverse topics and brands across different social media platforms (Xu et al., 2022). Applications of sentiment analysis in social media are detailed in Table 10.

3.8. Product reviews and consumer feedback

Utilizing sentiment analysis on product reviews and consumer feedback automates the identification of overall sentiments expressed by customers, categorizing them as positive, negative, or neutral. This application aids businesses in acquiring insights into customer satisfaction, pinpointing areas for enhancement, and making informed, data-driven decisions to elevate product quality and improve the overall customer experience (Ganesan et al., 2023). It is also valuable for tracking trends and monitoring brand reputation in real time. Table 11 discusses some recent applications of sentiment analysis in product reviews and consumer feedback.

Table 6
Summary of applications of sentiment analysis in customer support.

Ref.	Description
Storey and Park (2022)	Used an ontology to analyze customer support forum posts from large companies, offering insights into emotion generation in these interactions.
Jain (2021a)	Demonstrated the efficacy of a weak supervision approach combining weak sentiment classifiers by providing insights into the connection between customer sentiment and problem resolution in chat interactions.
Carvalho et al. (2023)	Demonstrated the effectiveness of BERT-based classifiers in sentiment analysis within Portuguese customer support conversations, highlighting their relevance in real-world applications.
El-Ansari and Beni-Hssane (2023)	Created a tailored chatbot for customer service in e-commerce applications. Using sentiment analysis and personalization techniques enhanced the effectiveness and efficiency of question-answering systems and chatbots.
Borg and Boldt (2020)	Developed a tailored sentiment analysis framework for Swedish Telecom's customer emails, amalgamating VADER sentiment, a Swedish lexicon, and SVM models. Enabled proactive customer support strategies through robust sentiment extraction and predictive capabilities within email threads
Capuano et al. (2021)	Proposed a Hierarchical Attention Networks-based CRM sentiment analysis, evolving via CRM operator feedback. Demonstrated superior classifier accuracy using a dataset and showcased an adaptive retraining mechanism that enhances performance without degradation.
Jain (2021b)	Used pre-trained model to solve weak sentiment classifiers and domain-specific rules to train a client conversation sentiment classifier with weak supervision, outperforming the Google Cloud NLP API in domain-specific cases.

Table 7
Summary of applications of sentiment analysis in entertainment.

Ref.	Description
Modi et al. (2022)	Using the Flask environment offered web-based data visualization. It also included language and comparative sentiment analysis for multinational companies and exploring interpretation methods for Twitter data.
Ghosal and Jain (2023)	Introduced an innovative multilevel knowledge representation framework that leverages Word2Vec operators and extended Ordered Weighted Average (OWA) operators for sentiment analysis in entertainment.
Sultana et al. (2022)	Developed a Bangla Aspect-Based Sentiment Analysis model specifically for the entertainment domain, which involved the collection of 4012 Bangla text comments from YouTube related to various entertainment topics.
Tunca et al. (2023)	Presented an innovative approach employing computational thematic content analysis to delve into the metaverse concept, revealing its key themes, sentiment distribution, and implications for entertainment technology.
Kumar et al. (2020)	"Developed a hybrid recommendation system merging collaborative filtering, content-based filtering, and sentiment analysis of movie tweets. Showcased trends, public sentiment, and user response to movies, validating the model's effectiveness through experiments on a public database".
AlSulaim and Qamar (2021)	Constructed an effective sentiment analysis model using trained on a substantial MyAnimeList dataset. The study aids decision-making in the entertainment sector and supports Anime producers in understanding viewer sentiments and satisfaction levels.
Model (2023)	Utilized pre-trained deep learning models, such as GPT-2 and CLIP, to introduce Gennie, a pioneering system for collaborative storyboard co-creation with users. The collaborative process is visualized, and the system's outputs undergo evaluation through a combination of qualitative and quantitative methods, complemented by user surveys.

4. Sentiment analysis through computational techniques

Sentiment analysis is not limited to text alone; incorporating multimodal approaches allows us to consider diverse sources of information, such as images, audio, and video, which can provide a more comprehensive understanding of sentiment. Multimodal sentiment analysis is a computational technique that combines multiple sources of information, such as text and images, to understand and analyze users' emotional states. This approach goes beyond traditional sentiment analysis, focusing on predicting positive or negative sentiment in text alone. By incorporating visual analysis and natural language processing, multimodal sentiment analysis aims to infer the latent emotional state of users based on the emotion word tags they attach to their posts. The use of deep neural networks in this approach has shown promising results, outperforming separate models based solely on either images or text (Hu and Flaxman, 2018). Additionally, the Labels perspective in sentiment analysis involves considering the nuanced nature of sentiment labels and acknowledging the subjective and context-dependent nature of emotions. Understanding this perspective is crucial for developing more accurate, context-aware sentiment analysis models. This section will briefly discuss multimodal sentiment analysis and labels perspective in sentimental analysis.

4.1. Multimodal sentiment analysis

Multimodal emotion analysis involves examining textual, audio and visual data to discriminate and interpret emotions (Das and Singh, 2023). This integrated approach allows for a more nuanced understanding of emotion in various contexts, including verbal reviews, vlogs, human–machine interactions, and social media. By integrating information from multiple formats, researchers aim to capture the richness of human expression, considering not only spoken or written words but

Table 8
Summary of applications of sentiment analysis in education domain.

Ref.	Description
Dake and Gyimah (2023)	Utilized various machine learning classifiers to lead the exploration of qualitative feedback provided by students, extracting valuable insights from their open-ended responses within educational settings.
Usart et al. (2023)	Proposed a sentiment analysis approach that considers gender sensitivity to assess the emotional atmosphere among individuals undergoing teacher training in online courses.
Tubishat et al. (2023)	Collected and analyzed many tweets and provided valuable insights into public sentiment around ChatGPT in education.
Alassaf and Qamar (2022)	Used one-way analysis of variance combined with Naïve Bayes classifiers and Support Vector Machine to categorize the emotions expressed in Arabic tweets, especially those related to Qassim University in Saudi Arabia, and achieved improved results.
Siddique and Kumar (2023)	Developed a custom sentiment analysis method called Senti_NEP using machine learning approaches for analyzing tweets related to education and the National Education Policy-2020 (NEP) in India.
Zhai et al. (2020)	Introduced Multi-AFM, merging global and local attention with gating units for enhanced contextual representation. Included a novel local attention method using dependency trees, outperforming existing models across diverse domains.
Liu et al. (2023)	Utilized multi-dataset pre-training, incorporating journal articles, a substantial student response dataset, and a focused scientific argumentation dataset, to enhance BERT and SciBERT models, ultimately improving automatic scoring in science education assessments.

Table 9
Summary of applications of sentiment analysis in e-commerce.

Ref.	Description
Yin et al. (2022)	Applied sentiment analysis to Twitter data from Lazada and Shopee, offering industry-specific insights and practical implications for e-commerce companies based on public sentiment.
Jha et al. (2021)	Proposed the use of PySpark and the flexible distributed dataset (RDD) for sentiment analysis, employed Restful APIs based on FLASK and Scrapy for data collection through web scraping, and presented practical applications of sentiment analysis for real-time product recommendations and insights derived from customer reviews and purchase behavior.
Karn et al. (2023)	Presented the Hybrid Recommendation Model (HRM) combined with hybrid sentiment analysis, effectively improving recommendation systems' accuracy and correctness.
Kaur and Sharma (2023)	To provide businesses with a means to analyze social sentiment around their products or services, introduced a powerful hybrid approach for sentiment analysis, effectively addressing ambiguity and inconsistency in sentences.
El-Ansari and Beni-Hssane (2023)	Introduced a personalized chatbot for e-commerce that incorporates sentiment analysis, ultimately providing tailored user experiences and addressing gaps in existing research.
Demircan et al. (2021)	Developed Turkish sentiment analysis models categorizing product reviews into positive, negative, and neutral sentiments. Ranked methods for decomposing sentiments in review texts and offered insights into social media sentiments using supervised machine learning.
Zhang et al. (2020a)	Used a directed weighted model for precise sentiment multi-classification in reviews, offering efficient classification under limited threshold rules. Provided a versatile framework for studying this approach in sentiment analysis.

also facial expressions, tone of voice, and visual content. The challenges lie in merging features and applying advanced techniques, such as deep learning, to analyze and interpret emotion in various forms of communication effectively.

4.1.1. *Image–text approach*

The "image-text approach" involves analyzing visual content and accompanying textual information to comprehensively understand emotion in multimedia data. A joint image-text sentiment classifier uses a new feature vector obtained by employing a fusion technique from both image and text data to predict sentiment (Das and Singh, 2023). A combination of visual–textual sentiment framework named multimodal attentive fusion was suggested by Huang et al. (2019). First, two separate attention models were introduced to learn the text and image models individually, named the semantic and visual attention frameworks. After that, a multimodal attention model was explored to extract the correlation between textual and visual features. Finally, a decision level was employed to integrate all three attention models to predict the sentiment polarity. Wang et al. (2014) proposed a joint image-text framework for the microblog images using a cross-media bag of words approach. They collected a total of 5000 microblog images from the Sina Weibo website. After employing a feature-level fusion technique, logistic regression, SVM and naïve Bayes machine learning classifiers were used to train the proposed system. The experimental results reveal that the joint image-text model outperformed the text-based model

4.1.2. *Audio-visual approach*

Audio–video multimodal sentiment analysis shows that audiovisual fusion mainly focuses on human faces. An audiovisual sentiment analysis framework was proposed by Yadav et al. (2015) by extracting emotion-related information from both the video and the audio channel

Table 10
Summary of applications of sentiment analysis in social media.

Ref.	Description
Benrouba and Boudour (2023)	Proposed and implemented an approach to filter emotionally harmful social media content for user well-being, utilizing Twitter data and natural language understanding to classify emotions, thereby enhancing the quality of displayed content.
Mehmood et al. (2022)	Presented a novel framework that combines statistical and sentiment analysis, allowing universities to identify the topics students are discussing and areas for improvement on social media.
Jain et al. (2023)	Used VADER and LIME for enhancing interpretability in sentiment analysis of social media texts. It implemented these methods on data from platforms like Twitter, Facebook, and Reddit, effectively visualizing and explaining sentiment analysis results, ultimately surpassing contemporary systems in terms of interpretability.
Halawani et al. (2023)	Proposed a model that automates social media sentiment analysis using Harris Hawks Optimization and deep learning. It reduced language processing dependency on data pre-processing through fastText-based word embeddings and skip-gram methods.
Asif et al. (2020)	Conducted sentiment analysis on multilingual social media for extremism intensity using a validated lexicon. Employed Multinomial Naïve Bayes and Linear Support Vector Classifier models, achieving 82% accuracy on Facebook text to identify extremist sentiments.
Lamsal et al. (2023)	Introduced CrisisTransformers, an ensemble of pre-trained language models and sentence encoders, leveraging transformer-based pre-training on a vast corpus of crisis-related social media texts for tasks such as classification, semantic search, and clustering in crisis informatics.
Kumar et al. (2023)	Used pre-trained Generative Adversarial Networks, specifically Deep Convolution-based GAN models, assessing their performance with Inception Score and Fréchet Inception Distance. It also delved into manipulation and detection techniques and anticipated future trends in the deepfake domain.

Table 11
Summary of applications of sentiment analysis in product reviews and consumer feedback.

Ref.	Description
Huang et al. (2022b)	Introduced the ERNIE-BiLSTM-Att (EBLA) emotion analysis model, which effectively combines ERNIE word embedding, bidirectional long-term, short-term memory and an attention mechanism to enhance sentiment analysis in Chinese e-commerce product reviews.
Huang et al. (2022b)	Proposed a novel approach predicting the usefulness of reviews based on sentiment analysis, emphasizing sentiment characteristics such as positivity, negativity, and emotions extracted from a sentiment dictionary.
Yadav (2023)	Introduced a new approach for personalized product recommendations, leveraging sentiment analysis on customer feedback and using the WordNet database to extract aspects and related emotions.
Deepa et al. (2023)	Developed the Novel Convolution Neural Network (N-CNN) system to understand customer emotions through sentiment analysis better.
Karn et al. (2023)	Proposed a new recommendation model that combines a hybrid recommendation model with hybrid sentiment analysis to enhance the accuracy and correctness of recommendation systems (RS).
Yang et al. (2020)	Introduced SLCABG, a novel sentiment analysis model merging lexicons and CNN-BiGRU. Tested on real book reviews, analyzing performance factors and evaluating via standard metrics for effectiveness.
Zhou et al. (2020)	Proposed SentiX, a pre-trained model on extensive review datasets, which excels in cross-domain sentiment analysis without fine-tuning, outperforming BERT with just 1% of the training samples.
Mandvikar (2023)	Used LLMs in AWS's Intelligent Document Processing workflow, highlighting their benefits in document classification, extraction, review, enrichment, and data integration while addressing challenges such as overfitting and computational costs.

in audiovisual content. For visual analysis, they extracted the facial expression features. The audio features such as pitch, pause, loudness and voice intensity were extracted from audio data. After integrating the feature vector, they predicted the sentiment polarity of the overall review according to the detected emotions. Chu and Roy (2017) proposed an audiovisual analysis method to generate emotional arcs for movies, including short videos on the web. They trained audio and video sentiment analysis models and then used them to construct separate emotional arcs for audio and visual content. They also conducted experiments to evaluate the micro-level performance and synthesize the audio and visual content prediction results.

4.1.3. Audio-image-text approach

The "audio-image-text approach" to multimodal emotion analysis integrates audio, image and text data to provide a holistic understanding of emotion across different sources of information. Poria et al. (2016) designed a real-time multimodal sentiment understanding model to harvest sentiments from web videos. They used feature and decision-level fusion methods to merge adequate information extracted from multiple modalities. The audio, video and textual features are merged using decision-level and feature-level fusion. The text sentiment analysis model has been enriched by semantic computing-based features, which significantly improved the performance of the model. Cambria et al. (2018) proposed a deep neural framework for multimodal sentiment analysis. They extracted the textual features using a convolutional neural network and the OpenSMILE software for audio features. The visual features are extracted using a deep convolutional neural network. All the feature vectors were combined for sentiment classification using support vector machine classifiers.

4.2. Labels perspective in sentimental analysis

The labels perspective in sentiment analysis refers to categorizing or annotating data points with sentiment labels, such as positive, negative or neutral, for training and evaluating sentiment analysis models. This perspective is crucial for understanding how well models interpret and attribute emotion to textual, audio, or visual content, contributing to sentiment analysis systems' overall effectiveness and accuracy.

4.2.1. Document level sentimental analysis

Document-level sentiment analysis involves evaluating sentiment in individual documents, such as single reviews. Challenges include irrelevant sentences, necessitating subjectivity/objectivity classification. Both supervised and unsupervised learning methods can be employed, using features like term frequency and opinion words (Balaji et al., 2017).

4.2.2. Sentence level sentiment analysis

Sentiment analysis at the sentence level focuses on evaluating emotion in individual sentences (Araújo et al., 2020). This approach allows for a detailed examination of the views expressed in short text sections, helping to understand the nuances of emotion. It is commonly used in scenarios where the polarity of emotion may vary within a document or conversation, offering a more detailed perspective of the emotional tone conveyed at the sentence level.

4.2.3. Phrase level sentiment analysis

Sentiment analysis at the phrase level involves the evaluation of the emotion in specific phrases or expressions. This fine-grained analysis allows for a more detailed understanding of variations in emotion within sentences, capturing nuances that may not be apparent at the sentence or document level. It is particularly useful in contexts where emotions can shift within a sentence, providing a more nuanced interpretation of the emotional tone associated with individual sentences.

Phrase-level sentiment analysis is one of the levels at which sentiment analysis is carried out. Sentiment analysis can be done at the word or phrase, sentence, and document levels. Different approaches and algorithms are used to realize sentiment classification problems, including lexicon-based, learning-based, and hybrid-based approaches (Memiş et al., 2024).

4.2.4. Word level emotion distribution learning

Word-level emotion distribution learning calculates the probability of associating a word with different emotions based on the intensity of the emotion, using the Valence-Arousal-Dominance (VAD) space. This method provides a broad word-emotion association by explicitly measuring the intensity of the emotion. Challenges include dealing with non-dictionary words, mitigated by using WordNet to extend the coverage of VAD lexical knowledge and handling non-emotional words or spelling errors (Li et al., 2021).

5. Datasets and pre-processing methods

This section will analyze the widely used Datasets and Preprocessing methods used in ML, DL, LLMs and Pre-trained model-based sentiment analysis.

5.1. Datasets

In sentimental analysis, datasets serve as the fundamental building blocks on which original research and valuable applications are based. Internet Movie Database (IMDb) (Maas et al., 2011), MovieLens (Singh et al., 2023), Amazon product reviews (Ghosh et al., 2016), Stanford Sentiment Treebank (SSTb) (Socher et al., 2013), CMU-MOSI (Zadeh et al., 2016), MOUD (Pérez-Rosas et al., 2013), Getty Images dataset (You et al., 2016), DengAI (Gupta et al., 2023), etc. datasets are used frequently to identify sentimental analysis. Table 12 provides an overview of a few popular datasets for sentimental analysis.

5.2. Pre-processing methods

In sentiment analysis, pre-processing is essential for transforming raw text data into machine-readable format. Methods such as data cleaning, tokenization, and missing data handling reduce noise, ultimately improving model accuracy in sentiment analysis. Table 13 discusses some prevalent pre-processing techniques.

6. Algorithms

Sentiment analysis, a core part of NLP, employs both ML and DL techniques. ML analyzes text using labeled data and features like word frequencies, while DL excels in complex pattern recognition. ML is interpretable and suitable for small datasets, while DL handles context effectively. The choice depends on the application and data size, from social media monitoring to customer feedback analysis. This section will briefly discuss some of the most used Machine Learning and Deep Learning algorithms for sentiment analysis. Fig. 5 briefly categorized the Machine Learning and Deep Learning algorithms used in sentiment analysis.

6.1. ML algorithms

Machine learning algorithms in sentiment analysis employ computational models to analyze and classify text-based data, discerning sentiments as positive, negative, or neutral. Through techniques like natural language processing and supervised learning, these algorithms provide automated insights into the emotional tone of textual content. This section will discuss some commonly used ML techniques in sentiment analysis.

6.1.1. Support vector machine (SVM)

The SVM is a powerful supervised learning method for regression and classification (Hearst et al., 1998). It comprises the Support Vector Regression (SVR) for regression tasks and the Support Vector Classification (SVC) for classification. In type, SVC determines the optimal hyperplane for binary and multi-class scenarios, creating a linear or non-linear boundary in the input space. Often associated with Support Vectors, Kernels are employed to define the separation function. The formulation of SVM is given below:

$$f(x) = \sum_{x_j \in S} {}_j y_j K\left(x_j, x\right) + b \tag{1}$$

Here in Eq. (1), x_i signifies training patterns, y_i represents class labels ($y_i \in \{+1, -1\}$), and S is set of Support Vectors. The dual formulation leads to a minimization problem expressed as:

$$\min_{0 \le i \le C} W = \frac{1}{2} \sum_{i,j} Q_{ijj} - \sum_i +b \sum_i y_{ii}. \tag{2}$$

In Eq. (2), $Q_{ij} = y_i y_j K(x_i, x_j)$ represents symmetric positive kernel matrix. C penalizes error values, and the conditions for dual are:

Table 12
Overview of popular sentiment analysis datasets.

Ref.	Dataset	Type	Domain	Feature
Xu et al. (2019)	15000 hotel comment texts	Text	E-commerce	The dataset includes 15000 positive and negative comments.
Ouyang et al. (2015)	Movie review excerpts (rotten-tomatoes.com)	Text	Entertainment	The dataset comprises 10,662 movie review sentences, evenly divided between positive and negative sentiments.
Singh et al. (2023)	MovieLens	Text	Entertainment	The dataset consists of approximately 100,000 ratings, spanning a scale of 1 to 5, provided by 943 individuals for 1,682 films.
Socher et al. (2013)	SSTb	Text	Entertainment	The dataset comprises 11,855 movie reviews, comprising 10,662 sentences evenly divided between positive and negative sentiments.
Maas et al. (2011)	Internet Movie Database (IMDb)	Text	Entertainment	The dataset is an extensive and balanced compilation of 50,000 movie reviews. It has been evenly divided with 25,000 reviews into two sets. Half of the reviews in each set are labeled as positive and negative. The dataset includes two columns: "review" and "sentiment".
Rane and Kumar (2018)	Twitter US Airline Sentiment	Text	Customer Support	The dataset comprises positive, negative, and neutral sentiments, with sample sizes of 2,363, 9,178, and 3,099, respectively.
Zadeh et al. (2016)	CMU-MOSI	Video	Social Media	The multimodal dataset includes 2,199 opinionated statements collected from 93 videos on diverse topics sourced from YouTube.
Pérez-Rosas et al. (2013)	MOUD	Video	Social Media	The multimodal dataset includes 79 videos focused on product reviews in the Spanish language.
You et al. (2016)	Getty Images dataset	Image, Text	Brand Monitoring	The dataset comprises 588,221 labeled instances, encompassing images and text.
You et al. (2015)	Twitter image dataset	Image	Social Media	The dataset includes 1,269 image tweets labeled by five Amazon Mechanical Turk (AMT) workers.
Gupta et al. (2023)	DengAI	Text	Healthcare	The dataset comprises approximately 1,450 entries documenting climate conditions and reported cases of dengue in two cities: San Juan and Iquitos.
Memon and Carley (2020)	COVID-19 Datasets	Text	Healthcare	The Twitter search API was employed with various keywords to gather data on three specific dates: March 29, 2020, June 15, 2020, and June 24, 2020.
Mohammed and Kora (2019)	Arabic-Egyptian corpus	Text	Entertainment	The dataset comprises 50,000 tweets written in informal Arabic, covering various topics.
Huang et al. (2022b)	Jingdong e-commerce	Text	Customer Support	The dataset encompasses 11,000 reviews, distributed as 6,000 positive reviews, 2,500 neutral reviews, and 2,500 negative reviews.
Pavitha et al. (2022a)	Reviews dataset	Text	Entertainment	The dataset consists of two columns: ' reviews' and 'comments'. It includes 3,943 positive comments and 2,975 negative comments.
Tang et al. (2015)	RT05	Text	Entertainment	The dataset comprises meetings involving 4 to 10 speakers and diverse acoustic conditions.
Alamoudi and Alghamdi (2021)	Yelp	Text, Image	Customer Support	The dataset is provided in JSON files, housing a total of 5,996,996 reviews, 188,593 businesses, and 280,992 pictures.
Aziz et al. (2023)	MSED	Text, Image	Social Media	The dataset comprises 9,190 instances of diverse, multimodal content, combining text-image data, spanning various cultures and languages, facilitating research into the interconnected nature of human desire, sentiment, and emotion within social media contexts.
Wu et al. (2022)	CH-SIMS	Video	Customer Support	Utilizing the CH-SIMS dataset, a video-based multimodel dataset in Chinese, the paper addresses the challenge of exploring connections and extracting complementary information in video-based sentiment analysis alongside three English datasets.
Jbene et al. (2022)	MELD	Text	Social Media	For textual sentiment analysis using the MELD dataset, the nlpaug library's ContextualWordEmbsAug, altering words with similar ones at a 20% probability, proved most effective for data augmentation.
Zhang et al. (2020b)	ScenarioSA	Text	Social Media	ScenarioSA tracks evolving sentiments for each speaker in 2,214 conversations, crucial for understanding emotional dynamics and enhancing interactive sentiment analysis models' depth and accuracy.
Zhang et al. (2023b)	CMMA	Text, Video	Entertainment	The CMMA dataset, with 21,795 multi-modal utterances, serves as a benchmark for training and evaluating models in multi-affection conversational analysis, providing labeled data for sentiment, emotion, sarcasm, and humor across text, visuals, and acoustics.

$$g_i = \frac{\partial W}{\partial_i} = \sum_i Q_{ijj} + y_i b - 1 = y_i f\left(x_i\right) - 1 \tag{3}$$

$$\frac{\partial W}{\partial b} = \sum_j y_j = 0 \tag{4}$$

This categorizes the training set into the Support Vector set (0 <_i < C, g_i = 0), the well-classified set ($_i = 0, g_i > 0$) and the error set ($_i = C, g_i < 0$). Now, introducing a quadratic penalty factor C_0 for the error points results in a modified kernel function, which turns the problem into a linearly separable case:

$$K_0(x_i, x_j) = K(x_i, x_j) + \frac{1}{C_0}\delta_{ij} \tag{5}$$

6.1.2. Naïve Bayes

Bayesian network classifiers (Friedman et al., 1997) are extensively employed in supervised classification, with the Naïve Bayes classifier being a notable example (Rish et al., 2001). This probabilistic model, rooted in Bayes' theorem, relies on the simplifying assumption of solid independence. Initially recognized in text retrieval, it continues to serve

Table 13
The table discusses the commonly used pre-processing methods in sentiment analysis.

Pre-Processing method	Description
Data cleaning	Data cleaning is the process of identifying and correcting errors, inconsistencies, or inaccuracies in datasets to enhance their quality and reliability for analysis (Rahm et al., 2000).
Tokenization	Tokenization breaks text into individual words or tokens, a fundamental step in natural language processing and text analysis (Aliwy, 2012).
Feature selection	Feature selection aim to enhance model performance and reduce dimensionality by selecting the most pertinent and informative features from a dataset (Guyon and Elisseeff, 2003).
Feature extraction	Feature extraction involves transforming data into a new representation, capturing essential information for analysis while reducing dimensionality (Guyon and Elisseeff, 2006).
Normalization	Normalization, in the context of machine learning, involve adjusting numerical data to a common range, typically between 0 and 1, to standardize measurements and facilitate reliable comparisons (Patro and Sahu, 2015).
Standardization	Standardization involves transforming numerical features with a mean of 0 and a standard deviation of 1. It ensures the features follow a standard normal distribution, making them comparable and aiding machine learning model performance (Ali et al., 2014).
Noise elimination	Noise elimination aims to reduce or remove irrelevant or random fluctuations in data to enhance the signal-to-noise ratio, improving the quality of information extracted for analysis or modeling (Gamberger et al., 2000).
Data augmentation	Data augmentation is a technique that enhances the size and diversity of a dataset by applying various transformations (e.g., rotation, cropping) to the existing data, aiding in better model training and generalization (Maharana et al., 2022).
Missing data handling	Missing data handling involves strategies for managing and filling in missing values in a dataset, ensuring that the data remains usable for analysis and modeling (Joel et al., 2022).
Stemming	Stemming is a text normalization technique that reduces words to their root or base form by removing suffixes, helping in text analysis and information retrieval (Gupta and Arora, 2022).

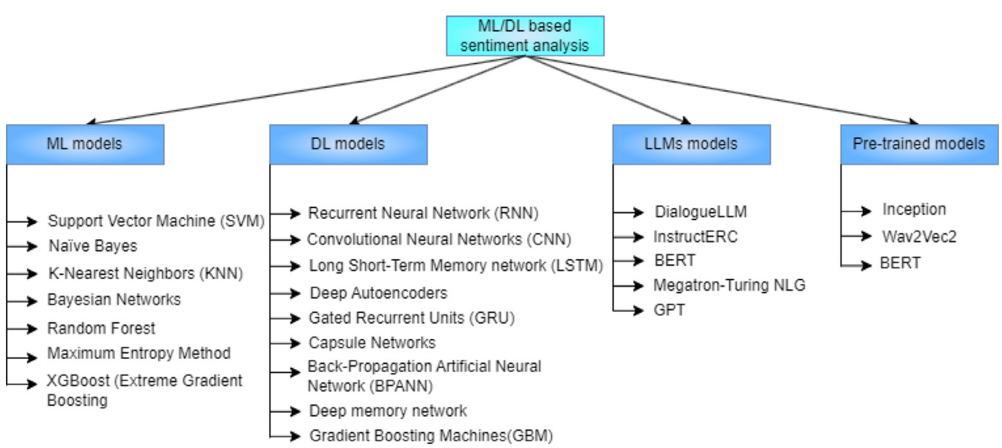

Fig. 5. This figure illustrates the taxonomy of machine learning and deep learning algorithms used for sentiment analysis.

as a fundamental method in text categorization. The primary objective in this context is assigning documents to categories based on word frequencies. The Bayesian classification approach offers practical learning algorithms and the ability to combine prior knowledge with observed data. In Naïve Bayes, the fundamental concept involves computing category probabilities for a given text document by evaluating the joint probabilities of words and categories, assuming word independence. The foundation lies in Bayes' theorem for conditional probability, stating that, for a given data point x and class y:

$$P(y/x) = \mathrm{P}(x/y)/\mathrm{P}(x) \tag{6}$$

Moreover, assuming that for a given data point, x = x1 to xi, Each attribute's likelihood of occurrence in a specific class is considered independent. Hence, the probability of x can be approximated as follows:

$$\mathrm{P}(y/x) = \mathrm{P}(y) \cdot \Pi \mathrm{P}\left(x_i/c\right) \tag{7}$$

6.1.3. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm is known for its excellent performance in classification and regression tasks, especially when dealing with multimodal inputs (Peterson, 2009). It operates by using feature vectors from training data to classify images effectively. KNN can be optimized for computational efficiency by either reducing the size of the training set or employing faster KNN variants. The choice of K-neighbors is determined by distance metrics, followed by a voting process to group data points. KNN is particularly suitable for classifying images that rely on local features. The algorithm's efficacy relies on choosing suitable distance metrics, with Euclidean Distance being a commonly used metric, along with others like Cosine, Minkowsky, and Chi-Square.

The widely used Euclidean distance calculation involves two feature vectors X = (x1 to xn) and Y = (y1 to yn) in an n-dimensional space. It is defined as:

$$Dist(E)(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}. \tag{8}$$

The Cosine distance, which focuses on element homogeneity, can be calculated using:

$$Dist(c)(X,Y) = \frac{x \cdot y}{\|x\|\|y\|}. \tag{9}$$

Another prevalent metric, which is the Minkowsky distance, is computed as:

$$Dist(m)(X,Y) = \left(\sum_{i=1}^{n}|x_i - y_i|^m\right)^{\frac{1}{m}}. \tag{10}$$

When $m = 1$, it denotes the Manhattan Distance, and $m = 2$ it signifies the Euclidean Distance. While 'm' can take any value, it is typically set to 1 or 2.

6.1.4. Random forest

Random Forest is a supervised ML method for classification and regression tasks. The approach involves training an extensive set of decision trees and aggregating their predictions to produce an outcome. This algorithm incorporates concepts like random subspaces and "bagging", it was formally introduced by Breiman (2001). The algorithm is trained using multiple decision trees, each fed with slightly varied data samples. Random Forest is a collection of algorithms utilizing decision trees as individual predictors, incorporating bagging, randomizing outputs, and Random Subspace, including Boosting. Recognized as a highly impactful classification algorithm, it accurately categorizes large datasets. This ensemble learning technique, suitable for both regression and classification, generates multiple decision trees in the training phase and determines the class mode by aggregating the outputs from individual trees.

6.1.5. Maximum entropy method

The Max Entropy algorithm is a probabilistic classifier categorized within models utilizing either a variable in an exponent or an exponential function (Jaynes, 1982). It avoids prematurely assuming the independence of structures. MaxEnt operates on the "principle of maximum entropy", selecting the model with the highest entropy that aligns optimally with our 'training data' among all possibilities. Typically, the Max Entropy classifier addresses extensive challenges in text categorization, such as language identification, subject classification, sentiment analysis, and similar tasks.

The maximum entropy classifier in text classification utilizes interdependent features, guided by the principle of "model all that is known and assume nothing about that which is unknown". The approach involves identifying classifiers 'empirically compatible' with training data, selecting the one maximizing Entropy. The goal is to extract contextual and statistical information from the document, categorizing elements like unigrams, bigrams, and other textual components into specific classes, such as 'objective/subjective' or 'positive/neutral/negative'.

6.1.6. XGBoost (Extreme Gradient Boosting)

XGBoost, or Extreme Gradient Boosting, is categorized among machine learning algorithms that originate from ensemble methods relying on decision trees. It works in the context of Gradient Boosting, effectively leveraging decision trees to enhance predictive performance. In this context, "XGBoost" is short for "Extreme Gradient Boosting" (Chen et al., 2015).

Ensemble learning provides a method for leveraging the predictive abilities of multiple learners. In the boosting paradigm, trees are built sequentially, with each successive tree focused on minimizing the errors of the preceding one. Each tree learns from the previous ones and adapts by addressing the remaining errors. Consequently, the subsequent tree in the sequence learns from the adjusted residuals. Typically, reinforcement employs weak learners as base models, known for their high bias. These weak learners collaboratively provide valuable insights into predictions, allowing the boosting technique to create a resilient learner by efficiently merging the capabilities of these individual weak learners.

Assume a training dataset is denoted as x_t and their corresponding labels y_t. In the XGBoost algorithm, a classifier generates the final prediction $\hat{y}_l^t$, representing the predicted value or label for a given input x_t.

$$\hat{y}_l^t = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_t^{t-1} + f_t(x_i) \tag{11}$$

Where $\hat{Y}_t^{(t-1)}$ represents the prior prediction and $f_t(x_l)$ is denotes the new prediction.. The objective in XGBoost is to minimize the following objective function to attain a high-quality model.

$$\mathcal{L}^t = \sum_{l=1}^{n} l(y_i, \hat{y}_l) + \Omega(f_t) \tag{12}$$

Here, objective function incorporates loss function $l(y_i, \hat{y}_l)$ and regularization term $\Omega(f_t)$. With the existence of Eq. (11). Now, rewrite the objective function as follows.

$$\mathcal{L}^t = \sum_{l=1}^{n}\left(l(y_i, \hat{y}_l^{(t-1)}) + f_t(x_i)\right) + \Omega(f_t) \tag{13}$$

The loss function is a measure to evaluate how effectively the model aligns with the training data, while regularization evaluates the complexity of the decision trees. The optimization of the loss function seeks to generate predictive models with enhanced accuracy, whereas optimizing regularization promotes the development of simpler and more generalized models.

6.2. DL algorithms

Deep learning algorithms in sentiment analysis use neural networks with multiple layers to automatically learn intricate patterns and representations from textual data. These models excel at capturing complex relationships within language, allowing them to discern and classify nuanced sentiments in a more nuanced and accurate manner. Widely used DL techniques in sentiment analysis are discussed below.

6.2.1. Convolutional neural networks

Convolutional neural networks (CNN) are artificial neural networks widely utilized in tasks associated with image recognition and computer vision. Drawing inspiration from the human visual system, CNNs have the capability to learn and recognize patterns in images without pre-existing information autonomously (O'Shea and Nash, 2015).

CNNs are composed of three primary types of layers: Convolutional layers, pooling layers, and fully connected layers, as illustrated in Fig. 6. The input layer contains pixel values extracted from the image. Convolutional layers determine the output of neurons connected to localized portions of the input by calculating the scalar product between their weights and the corresponding input region. The pooling layer conducts spatial downsampling, further decreasing the number of parameters in the activation. Various pooling techniques have been devised for diverse CNN applications, such as Max pooling, average pooling, stochastic pooling, and spatial pooling units. Fully connected layers then aim to produce class scores for classification based on the

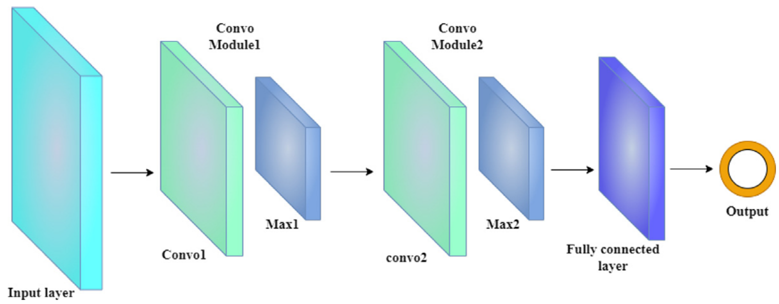

Fig. 6. Convolutional neural network.

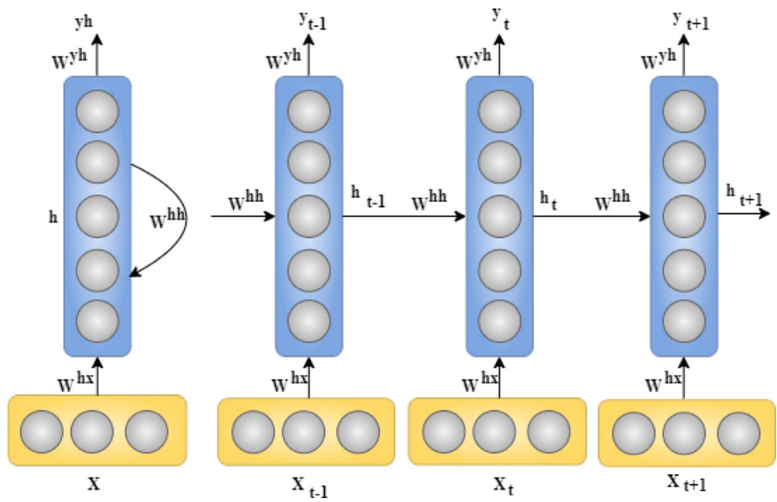

Fig. 7. Recurrent neural network.

activations. Incorporating Rectified Linear Units (ReLu) between these layers is often advised to enhance performance. The computation of a convolutional operation's outcome is expressed as:

$$C_i^{1,j} = \sigma(b_j^l + \sum_{m=1}^{M} W_m^{l,j} \; X_{i+m-1}^{l-1,j}) \tag{14}$$

In this equation, where l represents the layer index, σ signifies the activation function, b corresponds to the bias term for the feature map, M denotes the kernel/filter size, and W stands for the weight of the feature map, a Convolutional Neural Network (CNN) calculates activation unit values for various regions within the network. This process aids in the discovery of patterns across the input data.

6.2.2. Recurrent Neural Network

The Recurrent Neural Network (RNN) is a variant distinguished by interconnected neurons forming a loop structure. In contrast to standard feedforward neural networks, RNNs feature an inherent "memory" function, enabling them to handle sequences of inputs adeptly. This quality renders RNNs highly effective for tasks that entail sequential data processing (Medsker and Jain, 2001).

This concept of "memory" means that an RNN can perform a consistent operation for every element within a sequence. Crucially, every output generated by the RNN is influenced by all prior computations, effectively preserving information about the processing history, analogous to how humans recollect past information. Fig. 7, illustrates the architecture of RNN. On the left side, the figure presents an unfolded network featuring cyclic connections, while the right side portrays a simplified perspective of the same network presented as a sequence with three specific time steps. The count of time steps aligns with the length of the input sequence. For example, if the input is a sentence comprising six words, the RNN would be unfolded into a neural network with six time steps or layers, where each layer corresponds to one word.

In Fig. 7, x_t represents the input vector during a specific time step t. h_t denotes the hidden state during a specific time step t, calculated by considering the preceding hidden state during a specific time step. $t-1$ and the current time step's input t.

6.2.3. Gated Recurrent Units

The GRU (Gated Recurrent Unit) is a type of RNN architecture designed for sequential data processing and is an alternative to LSTM (Hochreiter and Schmidhuber, 1997). It addresses the vanishing gradient problem in traditional RNNs by incorporating gating mechanisms.

The main difference between GRU and LSTM lies in their internal architecture. While both are designed to capture long-term dependencies, GRU has a more straightforward structure with two (reset and update) gates, whereas LSTM has three (input, forget, and output). GRU is computationally less expensive and often requires fewer parameters than LSTM, making it more efficient in specific scenarios. However, LSTM may better capture more complex relationships in specific tasks.

6.2.4. Deep autoencoders

Autoencoders comprise an encoder and a decoder. While the encoder samples features from raw data, the decoder rebuilds the data using these extracted features (Bank et al., 2023). Deep autoencoders, an advanced version, employ deep learning techniques to enhance their performance (Kuchaiev and Ginsburg, 2017). Deep autoencoders create a compressed data representation by altering and reconstructing the input's dimensionality. They are versatile, learning compressed data encoding unsupervised and can be trained layer by layer, reducing computational requirements. Once well-trained, they extract crucial

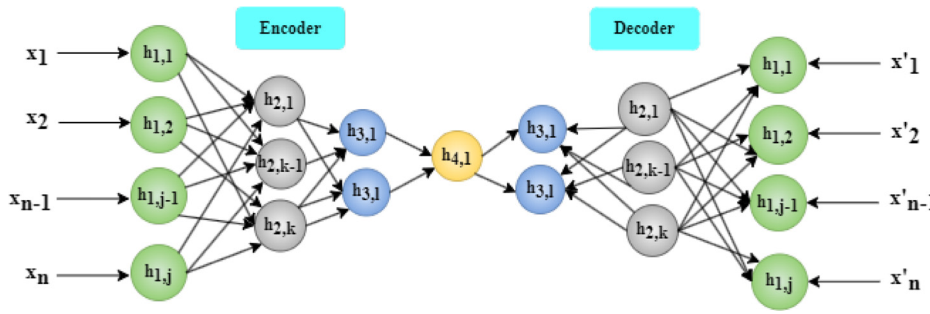

Fig. 8. Deep Autoencoders.

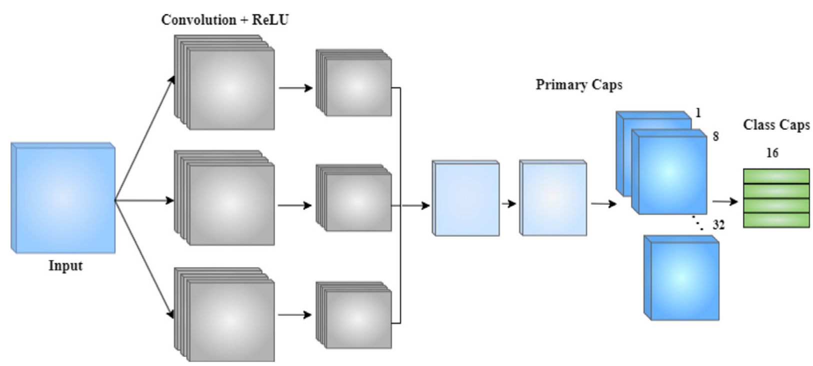

Fig. 9. CapsNet architecture for sentence classification.

features from input data. Since the hidden layers have lower dimensions than the input and output layers, the network serves as a tool for data encoding, effectively compressing features (as shown in Fig. 8).

In the standard deep autoencoder architecture, the neural network consists of distinct layers, each with specific functions:

- Input Layer: This layer serves as the entry point for the initial data, encompassing features extracted from sources like images or other data formats. The number of neurons in this layer is equivalent to the dimensionality of the input dataset.
- Hidden Layers: These intermediary layers, commonly called encoding layers, encompass fewer neurons than the input layer. Their objective is to establish a bottleneck in the network's topology, compelling it to acquire a concise, compressed representation of the input data.
- Output Layer: The ultimate layer, known as the output layer, is tasked with reconstructing the input data based on the encoded representation. Usually, it mirrors the number of neurons in the input layer, aiming to produce an output that accurately replicates the original input.

In Fig. 8, a distinguishing feature is that the hidden layers have lower dimensionality than the input and output layers, which characterizes it as an autoencoder. During training, the network minimizes the discrepancy between the input and output data, acquiring a concise representation within the hidden layers. The training process involves two stages: encoding, where data is transformed into a lower-dimensional representation, and decoding, which maps this representation back to the original data space. This dual process empowers deep autoencoders, with their multiple hidden layers, to capture hierarchical features and complex data structures, making them highly effective in tasks such as dimensionality reduction, data compression, and abstract feature extraction.

6.2.5. Capsule Networks

Capsule Network (CapsNet) introduced the notion of capsules, clusters of neurons where the activity vector denotes different parameters of a particular entity. The magnitude of a particular vector within a capsule indicates the likelihood of detecting a feature, with its orientation representing instantiation parameters. In CapsNet, the result from a specific capsule is conveyed to a suitable parent in the layer above using a mechanism referred to as dynamic routing (Patrick et al., 2022).

Besides retaining all spatial information, capsules demonstrate notably more excellent resistance to white-box adversarial attacks than conventional convolutional neural networks. The foundational architecture of CapsNet highlighted in Sabour et al. (2017) includes a convolutional layer, an additional convolutional layer known as the primary capsules layer, and the ultimate layer designated as the digital capsules layer, which is illustrated in Fig. 9.

6.3. Large Language Models

Large Language Models, exemplified by GPT, DialogueLLM, and InstructERC models, play a pivotal role in natural language processing. These models specialize in language tasks by leveraging extensive training data and advanced transformer architectures featuring over 100 billion parameters. Their proficiency extends across diverse language-related functions, including text generation, machine translation, question answering, summarization, content creation, and chatbot interactions. Prominent examples in the LLM landscape include GPT-3, LaMDA, Megatron-Turing NLG, Bloom, and WuDao 2.0.

Large Language Models have significantly reshaped the research climate of sentiment analysis by introducing a new paradigm that harnesses the power of deep learning and mass pre-learning on diverse textual data. LLMs, such as OpenAI's GPT-3, have demonstrated remarkable abilities in understanding context, creating coherent text, and capturing nuances of linguistic patterns. This transformative impact has led to a paradigm shift in sentiment analysis methodologies and applications.

Several studies have highlighted the importance of LLMs in sentiment analysis, emphasizing their effectiveness in capturing contextual information, understanding sarcasm, and dealing with complex linguistic structures. One such survey, conducted by Zhang et al. (2018), extensively discusses the role of LLMs in advancing sentiment analysis research. The survey underscores the ability of LLMs to outperform traditional approaches, especially when dealing with large and diverse datasets. Another notable survey by Chen et al. (2020) delves into the impact of LLMs on sentiment analysis tasks. The survey discusses the evolution from traditional methods to deep learning approaches, with a special focus on the contributions of LLMs in capturing semantic relationships and improving sentiment prediction accuracy. These surveys collectively highlight how LLMs have become pivotal in reshaping sentiment analysis research.

6.3.1. DialogueLLM

DialogueLLM is an open-source emotional Large Language Model specifically designed for Emotion Recognition in Conversations (ERC) tasks. It addresses the limitations of current emotion LLMs by incorporating context and emotion knowledge tuning and leveraging multi-modal information. DialogueLLM is trained using benchmarking multimodal emotional dialogs, where visual information is considered supplementary knowledge to construct high-quality instructions. The model undergoes an instruction-tuning step involving training on supervised instruction-input–output data to align with human prompts and enable precise customization for emotional domains (Zhang et al., 2023a). They assume that there are N conversation instances in the instruction dataset, where the ith conversation D_i contains K multi-modal utterances, represented as $D_i = \{(C_z, M_k), Y_k\}$, where C_z denotes previous z contextual utterances, M_k represents the kth target utterance to be classified, and Y_k signifies the emotion label of the kth target utterance. Here, $i \in [1, 2, \dots, N]$, $k \in [1, 2, \dots, K]$, and $z \geq 0$.

The target utterance consists of textual (T) and visual (V) modalities, i.e., $M_k = (T_k, V_k)$, where $T_k \in \mathbb{R}^{l_{T_k} \times d_{T_k}}$, $V_k \in \mathbb{R}^{l_{I_k} \times d_{I_k}}$. Here, l_{T_k} and l_{V_k} denote the sequence length of textual and visual utterances, and d_{T_k} and d_{V_k} represent the dimensions of the textual and visual features. After this, they summarize their research problem as follows: Given one multi-speaker conversation including K multi-modal utterances, how to detect their emotions? It could be written as:

$$\zeta = \prod_k p(Y_k | C_z, M_k, \Theta) \tag{15}$$

where Θ denotes the parameter set.

6.3.2. InstructERC

Reforming Emotion Recognition in Conversation with a Retrieval Multi-task LLMs Framework" by Lei et al. (2023). It aims to reformulate the emotion recognition in dialog (ERC) task from a discriminative framework to a generative framework based on Large Language Models.

Adopting the next token prediction loss measures the model's output error, forming the basis for the subsequent loss calculation. To sum up the instruction-based generative framework for ERC, given an input utterance x_i after concatenating the retrieval template $drvl$ and a Large Language Model (LLM), the model returns the logits g_i and the generated text y_i for the entire sentence, including both input and output tokens. The following equation represents this:

$$y_i, g_i = \text{LLM}(x_i, \theta) \tag{16}$$

Here, θ is the same as mentioned. The LLM predicts the conditional probability $p(\gamma_i | x_i, \theta)$ of generating each token γ_i of the generated text y_i until the end symbol $\langle eos \rangle$ is outputted. As for logits $g_i \in \mathbb{R}^{L \times V}$, where L and V denote the length of the entire sentence and the size of the vocabulary used by the LLM, respectively. In accordance with the original training method of LLMs, we adopt the next token prediction loss to measure the model's output error. Therefore, the loss calculation of the main task, denoted as L_{main}, is defined as follows:

$$L_{\text{main}} = \sum_i^N -\log P(\epsilon_i | x_i, \theta) \tag{17}$$

6.3.3. Bidirectional encoder representations from transformers (BERT)

Generic language models that have been pre-trained, such as BERT, have demonstrated exceptional performance across various natural language processing tasks (Alaparthi and Mishra, 2020). These models undergo unsupervised training on extensive text data and can be utilized for diverse tasks. BERT, in particular, has achieved considerable success and is constructed based on the Transformer architecture, which employs attention mechanisms in its encoder and decoder components.

It is crucial to highlight that BERT predominantly employs the transformer encoder and does not include a decoder network, as it is not tailored for sequence-to-sequence tasks. However, it remains adaptable for application in such contexts. Most language models utilize unidirectional architectures, meaning they generate outputs solely based on preceding words (the left context). This constraint becomes evident when employing these models for tasks where the complete text is accessible for prediction. BERT tackles this limitation by introducing a bidirectional language model architecture, enabling it to consider both the left and right context, which can be especially advantageous for tasks such as sentiment analysis.

6.3.4. Megatron-Turing NLG

Megatron-Turing NLG is a transformer-based language model with 530 billion parameters, making it the largest monolithic language model trained to date. It was developed through a collaboration between NVIDIA Megatron-LM and Microsoft DeepSpeed, using an efficient and scalable 3D parallel system that combines data, pipeline, and tensor-slicing-based parallelism. The training process involved building high-quality, natural language training corpora with hundreds of billions of tokens and developing training recipes to improve optimization efficiency and stability. MT-NLG has demonstrated superior zero-, one-, and few-shot learning accuracies on several NLP benchmarks, establishing new state-of-the-art results. Additionally, it has shown impressive qualitative performance in solving riddles, answering Jeopardy questions, and generating code (Smith et al., 2022; Shoeybi et al., 2019; Narayanan et al., 2021).

They used the architecture of the transformer decoder (Radford et al., 2019), which is a left-to-right, autoregressive, generative transformer-based language model, and scaled it up to 530 billion parameters. The number of layers, hidden dimensions, and attention heads are 105, 20480, and 128, respectively. The sequence length is 2048, and the global batch size is 1920. They employed 8-way tensor and 35-way pipeline parallelism. The learning rate is 5.0×10^{-5}. They utilized one billion tokens for linear learning rate warmup. Cosine decay was employed for the learning rate, targeting to reach 10% of its value over 340 billion tokens. Over the first 12 billion tokens, They started with a batch size of 32 and gradually increased it in increments of 32 until reaching the final batch size of 1920. Used the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.95$, and $\epsilon = 10^{-8}$. The gradient norm was clipped at 1.0, and a weight decay of 0.1 was applied. They used a normal distribution with zero mean and a standard deviation of 4.0×10^{-3} for weight initialization. Their training dataset consists of 339 billion tokens, and they trained MT-NLG on 270 billion tokens by blending the 15 training datasets as described above. They also set aside 2% of our data for validation.

A higher learning rate increases model instability. Used approximately $p^{1/(3 \times H)}$ as a standard deviation for weight initialization, where H denotes the size of the hidden dimension. Similar to Nguyen and Salazar (2019), they also observed that using higher variance for weight initialization fails to converge. They also reduced β_2 from its standard value of 0.99 to reduce spikes in the training loss.

6.3.5. GPT-3

GPT-3 has the ability to continue a prompt in the same style and content, and its use of style and content goes beyond syntax, as there is a bidirectional causation between the two. GPT-3's capacity to produce contextual first-person speech should not be confused with having a set

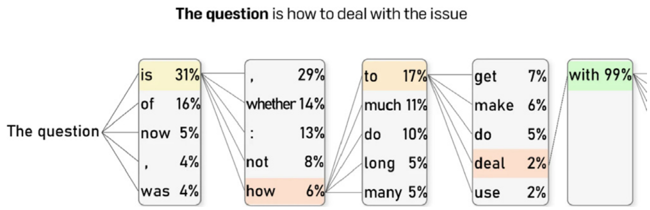

Fig. 10. Tree of possible continuations (Sobieszek and Price, 2022).

of views on any topic, as it is a language generator model. The discussion of answering questions in the context of computer intelligence, including GPT-3, has been a topic of debate for some time (Sobieszek and Price, 2022). Fig. 10 illustrates the diverse range of issues GPT-3 can address: Top: the prompt (the text provided continuation) is marked in bold. GPT's continuation is color-coded, representing the conditional probability of this continuation based on the previous text. Bottom: The probabilities over continuations, from which GPT picks with a weighted random draw. The probabilities in each step are determined by the choice in the previous step and the entire text before it. Note that if, at any step, the choice would have been different, the probabilities in every subsequent step would also be completely different.

GPT-3's implementation involves its ability to generate language-based responses, prioritize plausibility over truthfulness, and engage in various semantic tasks. However, it is important to note that GPT-3's responses should not be mistaken for having personal views or opinions (Floridi and Chiriatti, 2020).

6.4. Pre-trained

A pre-trained model is a machine learning model that has been trained on extensive datasets, acquiring a wealth of knowledge in tasks such as image recognition, natural language processing, speech recognition, and machine translation. These models serve as efficient starting points, saving time and resources by leveraging existing expertise. They can be employed directly for their intended purposes or fine-tuned for specific tasks, democratizing AI accessibility and fostering innovation across diverse domains. Examples include ResNet and VGG16 for image recognition, BERT and GPT-3 for natural language processing, and Wav2Vec2 for speech recognition. Overall, pre-trained models significantly accelerate development, enhance performance, and contribute to advancements in artificial intelligence.

6.4.1. Inception

An Inception network consists of modules of the above type stacked upon each other, with occasional max-pooling layers with stride 2 to halve the resolution of the grid. For technical reasons (memory efficiency during training), using Inception modules only at higher layers seemed beneficial while keeping the lower layers in the traditional convolutional fashion. This is not strictly necessary; it simply reflects some infrastructural inefficiencies in their current implementation.

A useful aspect of this inception model architecture Fig. 11 allows increasing the number of units at each stage significantly without an uncontrolled blow-up in computational complexity at later stages. Fig. 11(a) illustrates a basic module with 1 × 1, 3 × 3, and 5 × 5 convolutions and max pooling. 11(b) refines this with dimensionality reduction using 1 × 1 convolutions before 3 × 3 and 5 × 5 convolutions, enhancing computational efficiency. Both versions enable capturing features at different scales, which is crucial for processing visual information effectively. This is achieved by ubiquitous dimensionality reduction before expensive convolutions with larger patch sizes. Furthermore, the design follows the practical intuition that visual information should be processed at various scales and then aggregated so that the next stage can abstract features from the different scales simultaneously (Szegedy et al., 2015).

6.4.2. Wav2Vec2

Wav2vec2.0 is a model that has shown powerful representation ability and feasibility in ultra-low resource speech recognition tasks. It has been pre-trained on large amounts of unlabeled data using self-supervision and has been successfully applied to downstream tasks in the speech domain. The model has been examined on the Librispeech corpus, which belongs to the audiobook domain, but has not been tested on real spoken scenarios and languages other than English. However, when applied to low-resource speech recognition tasks in various spoken languages, it has achieved more than 20% relative improvements compared to previous work, with English achieving a gain of 52.4%. Coarse-grained modeling units, such as subwords or characters, have achieved better results than fine-grained units like phone or letter (Yi et al., 2020). In this Wav2Vec2 model architecture Fig. 12 where Left: the structure of wav2vec2.0 and corresponding self-training criterion. It contains a stack of convolution layers and self-attention layers., Right: two decoding branches that apply wav2vec2.0 to ASR tasks with additional projection or decoder, which is trained with CTC or cross-entropy loss, respectively.

7. State-of-the-art result analysis and their limitations

This overview explores the recent and significant progress in sentiment analysis, providing insights into state-of-the-art techniques, technologies, and research findings that have shaped the current sentiment analysis landscape. It offers valuable perspectives on the latest advancements, methodologies, and performance, providing researchers and professionals with insights into efforts to improve the accuracy and versatility of HAR methods. First, Table 14 examines recent state-of-the-art work in sentiment analysis utilizing machine learning models. Then, Table 15 discusses some DL-based state-of-the-art research experiments. After that, Table 16 and Table 17 address articles on LLMs and Pre-trained-based sentiment analysis, respectively.

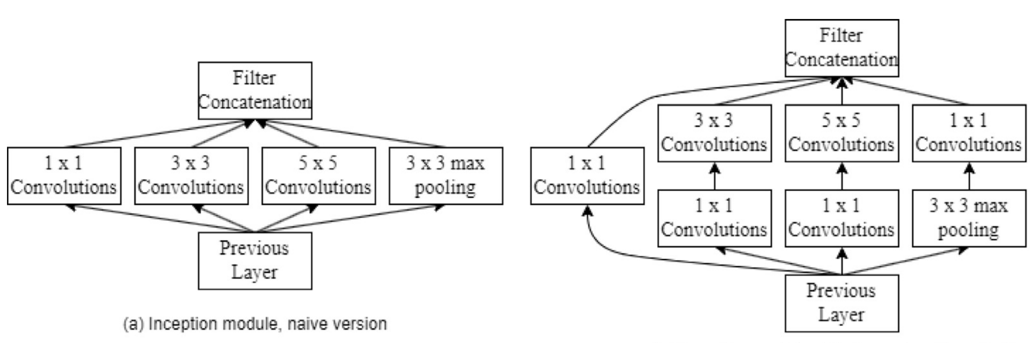

Fig. 11. Inception module.

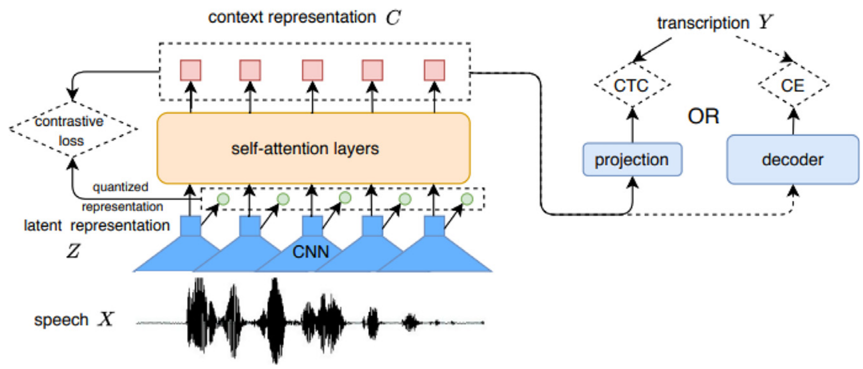

Fig. 12. Wav2Vec2 model overview (Yi et al., 2020).

8. Challenges and future research opportunities

Despite significant advances in sentiment analysis, numerous challenges persist, necessitating ongoing research efforts to address issues like bias, context understanding, and domain adaptation. Future research is crucial to enhance the robustness and applicability of sentiment analysis in diverse and evolving contexts. This section will discuss sentiment analysis's major and significant challenges and future research opportunities to overcome them.

8.1. Context understanding

Context Understanding is a fundamental aspect of sentiment analysis and entails grasping the intricate contextual nuances that influence the sentiment expressed in text. This poses a substantial challenge, as sentiments can be heavily reliant on context and can be influenced by elements such as sarcasm, irony, negation, word sense disambiguation, and ambiguity. For instance, in the sentence "The service at the restaurant was so slow that I had all the time in the world", the use of "all the time in the world" indicates sarcasm, where the intended sentiment is negative, despite the positive words used. Similarly, negation can reverse sentiment, as in "not bad", which actually conveys a positive sentiment. Word sense disambiguation is crucial in understanding words with multiple meanings, as in "bank" (financial institution vs. river bank). Ambiguity further complicates matters, as sentences with multiple interpretations can make it challenging to discern the intended sentiment. According to Zuo et al. (2020), distinguishing explicit from implicit emotions in text for detecting sentiment in sentiment analysis is challenging because explicit emotions are easily identifiable through sentiment words, while implicit emotions, without such words, remain elusive. Building sentiment lexicons necessitates substantial labeled data and resource-intensive feature selection, further complicating emotion interpretation in text. These complexities make context understanding a significant challenge in sentiment analysis, as overlooking these nuances can lead to erroneous interpretations.

Overcoming these challenges requires developing advanced algorithms for contextual analysis, improving word sense disambiguation techniques, and creating adaptable machine learning models. For instance, Bhatia et al. suggested a model that utilized a dynamic configuration window function that leverages the context around ambiguous terms to disambiguate word senses. This approach employs a genetic algorithm for optimization, incorporating linguistic knowledge from WordNet, resulting in an 8% accuracy improvement compared to existing methods, achieving an 80% accuracy rate (Bhatia et al., 2022). Furthermore, addressing text ambiguity and exploring sentiment analysis across modalities like text and images are essential. Tailoring models to specific domains and leveraging deep learning and NLP can advance sentiment analysis. Encouraging user-generated sentiment resources for comprehensive sentiment lexicons is crucial to reducing resource-intensive processes and enhancing sentiment interpretation in text.

8.2. Emojis and emoticons

Sentiment analysis encounters a notable challenge with emojis and emoticons, as they introduce added complexity to interpreting text-based emotions. These visual symbols can convey emotions that do not match the accompanying text (Shaik et al., 2023). For example, a sad emoji may contradict an "I'm fine" message. This challenge arises because of emojis' ambiguity and subjective interpretation, leading to possible misclassification by sentiment models (Venkataraman and Mohandoss, 2023). Challenges with emojis include the need for models to consider the nuanced emotional context introduced by these symbols, as misinterpretation can lead to inaccurate sentiment analysis, making it difficult to determine the true sentiment of the textual data. Grover et al. addressed that the ever-changing emoji vocabulary poses a challenge for sentiment analysis models, further complicated by the potential for emojis to be used sarcastically or ironically, leading to a disparity between their literal meaning and intended sentiment (Grover, 2022).

To overcome the challenges, sentiment models could incorporate contextual analysis techniques to discern the actual sentiment, accounting for situations where emojis contradict the accompanying text, such as recognizing sarcasm or irony. Furthermore, creating adaptable models that can evolve with the changing emoji vocabulary would be beneficial. For example, researchers can explore DL and NLP to build models capable of interpreting evolving emoji meanings as (Li et al., 2023) presented an innovative technique for emoji vectorization, generating emoji vectors that are subsequently incorporated into the sentiment analysis model.

8.3. Domain adaptation

Domain adaptation is a critical challenge in sentiment analysis due to the significant diversity in text data across domains, languages, and language styles. Sentiment analysis models often struggle with domain shifts, where distinct vocabularies, writing styles, and sentiment expressions in areas like product reviews, social media, or news articles impact their accuracy (Fang et al., 2022). Handling multilingual text is another complexity, as sentiments expressed in one language may not directly translate to another. Moreover, the ever-evolving nature of language, code-switching in multilingual contexts (Mabokela et al., 2022), sentiment analysis of code-mixed data (Shanmugavadivel et al., 2022), and limited resources for low-resource languages (Reitmaier et al., 2022) pose additional difficulties in adapting sentiment analysis models. Practical scenarios involve adapting models between domains like product reviews and social media, where informal language and unique expressions challenge model performance. Therefore, multilingual reviews requiring cultural and linguistic context, code-switching in customer service chats, and nuanced sentiment transitions across languages pose a significant challenge to domain adaptation.

Table 14
Results analysis of state-of-the-art research articles on ML-based sentimental analysis.

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Birjali et al. (2021)	Customer Support	Social media	Removing repeated characters, Features extraction, Features selection	SVM, ANN	Accuracy: 92.71%	**Scope Limitation:** The paper may overlook emerging sentiment analysis techniques due to its limited scope in addressing specific categories, hindering a complete understanding of the field. **Assessment Depth:** While briefly mentioning limitations, the paper lacks a deeper critical assessment of the strengths and weaknesses of discussed sentiment analysis approaches, potentially impacting researchers' decision-making.
Qorib et al. (2023)	Healcare, Education	Twitter COVID-19	Tokenization, Removing, Stemming, Lemmatization	TF-IDF + LinearSVC	Accuracy: 96.75%, Precision: 96.92%	**Limited Generalizability:** The study relies on social media data, specifically live-streaming public tweets, which may not represent the entire population's sentiments. The findings might be skewed towards the Twitter user demographic and may not generalize well to broader public attitudes. **Algorithmic Dependency:** The paper explores multiple sentiment computation methods and learning algorithms, but the choice of the best-performing model is based on experimental results.
Aslan et al. (2023)	Healthcare	Twitter API	Removing, Tokenization, Lemmatization, Stemming	Logistic regression, Decision tree	Accuracy: LR: 88.94%, DT: 92.53% F1-score: LR: 85.71%, DT: 90.54%	**Lack of Comparative Analysis:** The paper's failure to compare the TSA-CNN-AOA approach with existing sentiment analysis methods hinders a comprehensive assessment of its performance and effectiveness relative to state-of-the-art techniques. **Limited Transparency and Replicability:** The absence of a detailed explanation of the arithmetic optimization algorithm (AOA) used in the proposed approach makes it challenging for readers to understand and replicate the results, potentially limiting the credibility and reproducibility of the findings.
Dake and Gyimah (2023)	Education	Qualitative feedback from students	Tokenization, Data removing	SVM	Accuracy: 63.79%	**Limited Scope:** The paper's exclusive focus on sentiment analysis of qualitative feedback from students in a specific context restricts the generalizability of its findings to other feedback types or educational settings. **Algorithmic Narrowness:** The evaluation of only four classifiers (Naïve Bayes, Support Vector Machine, J48 Decision Tree, and Random Forest) without considering a broader range of machine learning algorithms limits the exploration of potentially more effective models for sentiment analysis in educational feedback scenarios.
Kheiri and Karimi (2023)	Product Reviews, Consumer Feedback	SemEval 2017	Removing noise, Tokenization	GPT 3.5 Turbo	Accuracy: 97.32%, Recall:91.98%, F1-score:94.26%	**Absence of Model Limitation Discussion:** The paper lacks a discussion on potential limitations or drawbacks of the proposed GPT-3.5 model, hindering a comprehensive understanding of its applicability and potential challenges. **Limited Comparative Analysis:** While the paper extensively explores the performance of GPT-3.5, it does not provide a comparative analysis with other large language models, such as BERT or RoBERTa, which could offer insights into the relative strengths and weaknesses of different state-of-the-art models in sentiment analysis tasks.
Shah et al. (2022)	Entertainment	Film review	Removing, Tokenization	KNN	Accuracy: 81.43%, Precision: 72.73%, Recall: 85.71%, F1-score: 78.69%	**Dataset Information Gap:** The paper lacks information regarding the size and diversity of the film review dataset, making it challenging to assess the dataset's representativeness and potential impact on the study's findings. **Omitted Classifier Analysis Details:** The paper does not delve into a comprehensive analysis of the K-nearest neighbors classifier's performance in comparison to the multinomial naive Bayes (MNB) classifier, leaving a gap in understanding the strengths and limitations of each approach.
Ahmed et al. (2023)	Customer Support	Twitter	Feature engineering, Feature grouping	C4.5 Decision Trees	Accuracy: 75% , F1-score: 75%	**Limited External Validity:** The study focuses on conversations between customers and a specific online retailer's agent (AmazonHelp) on Twitter, potentially limiting the generalizability of findings to other platforms, industries, or customer service contexts. **Algorithmic Performance Dependency:** The paper claims that decision trees outperformed other techniques without providing extensive comparative analysis. The limited discussion on performance comparison and the absence of benchmarking against state-of-the-art sentiment analysis methods raise concerns about the broader applicability and reliability of the proposed hybrid framework.

(continued on next page)

Table 14 (*continued*).

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Kabir et al. (2023)	E-commerce	Book reviews	Labeling, Translation	BERT	F1-score: 93.91%	**Limited Sentiment Nuance Representation:** The use of three broad sentiment categories (positive, negative, and neutral) in the dataset may oversimplify sentiment analysis, potentially missing out on more nuanced sentiments that could be crucial for a comprehensive understanding. **Bias in Dataset Source:** The dataset collected from two online bookshops introduces a potential bias towards book-related sentiment, limiting the generalizability of the findings to sentiments in other domains or contexts.

Table 15
Results analysis of state-of-the-art research articles on DL-based sentimental analysis.

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Vatambeti et al. (2023)	Consumer Feedback, E-commerce	Swiggy, Zomato, UberEats	Feature extraction	Conv-BiLSTM-EHO (Elephant herding optimization (EHO)	Accuracy: Swiggy: 98.08%, Zomato: 92.24%, UberEats: 93.4%	**Language Bias:** The study's analysis was confined to English tweets, potentially overlooking valuable data and sentiments expressed in other languages, limiting the study's cultural and linguistic inclusivity. **Omission of Non-Textual Elements:** Excluding emojis, slang, and non-English sentiment terms from the analysis narrows the scope of sentiment assessment, potentially missing nuances in user expression and reducing the overall comprehensiveness of the study.
Kaur and Sharma (2023)	Customer Support, E-commerce	STS-Gold	Feature extraction, Data cleaning, Tokenization	HFV+LSTM	Accuracy: 89.9%, F1-score: 91.3%	**Absence of Comparative Evaluation:** The paper lacks a comparative analysis with other existing models or methods, limiting the understanding of the proposed model's performance and effectiveness relative to alternative approaches. **Unexplored Real-world Implementation Challenges:** The paper misses discussing potential challenges or drawbacks in implementing the proposed model in real-world scenarios, leaving uncertainties about its practical applicability and potential limitations in varied environments.
Paramesha et al. (2023)	Entertainment	Movies	Feature extraction, Feature engineering, Lexicons	BERT	Accuracy: 96.3%, Precision: 96.3%, Recall: 96.3%	**Limited Generalization:** The paper primarily focuses on sentiment analysis in the context of product and service reviews, potentially limiting the generalizability of the proposed approaches to other domains or types of textual data where sentiment analysis might be applied. **Lack of Human Interpretability:** While machine learning and deep learning models are employed, the paper acknowledges that the results they produce may not be easily interpretable by humans. This lack of interpretability could challenge understanding and explaining the rationale behind the sentiment classifications, potentially affecting user trust and acceptance.
Vohra and Garg (2023)	Healthcare	Twitter	Data converting, Data cleaning, Tokenisation	CNN	Accuracy: 92.59%	**Limited Generalizability:** The study focuses on Twitter data related to the trend of working from home during a specific period. The temporal context may influence the findings and may not generalize well to different timeframes or platforms, limiting the broader applicability of the results. **Sentiment Annotation Dependency:** The sentiment classes are determined using VADER, a pre-trained sentiment analysis tool. While VADER is widely used, its accuracy is context-dependent, and the reliance on an external tool for sentiment annotation introduces a potential limitation in the study's independence and robustness.

(*continued on next page*)

In overcoming these challenges, exploring strategies to adapt sentiment analysis models to low-resource languages and enhancing performance in practical scenarios (Mamta et al., 2022), such as adapting between product reviews and social media, can be crucial. Incorporating cultural and linguistic context in multilingual sentiment analysis and refining models for nuanced sentiment transitions across languages are essential areas for future investigation. Research should aim to create more adaptable and accurate sentiment analysis models that can effectively handle the complexities introduced by diverse domains and linguistic variations.

8.4. Bias and Fairness

Bias and Fairness are substantial challenges in sentiment analysis. Bias refers to unfair, prejudiced, or skewed perspectives within sentiment models, leading to unequal treatment or misclassification of

Table 15 (*continued*).

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Kora and Mohammed (2023)	Product Reviews, Consumer Feedback	ASTD, ArSenTD-LEV, Movie Reviews, Twitter US	Data Cleaning	Meta-ensemble deep learning	Accuracy: ASTD: 77.6%, ArSenTD-LEV: 82.2%, Movie Reviews: 83.9%, Twitter US: 85.3%	**Complexity in Implementation:** The proposed meta-ensemble deep learning approach introduces a complex framework with three levels of meta-learners. While it aims to enhance sentiment analysis performance, its intricate structure may pose implementation and practical application challenges. **Limited Exploration of Meta-Predictions Impact:** While the paper discusses the impact of meta-predictions, it may benefit from a more in-depth exploration of how these predictions specifically contribute to the improved performance, providing deeper insights into the effectiveness of the proposed approach.
Hayawi et al. (2022)	Healthcare	ANTi-Vax	Stemming	BERT	Accuracy: 98%, F1-score: 98%, Precision: 97%, Recall: 98%	**Limited Generalization to Real-World Scenarios:** The study primarily focuses on sentiment analysis within specific domains, such as electronics, music, video games, books, and DVDs. The lack of diverse real-world domains may limit the generalizability of the proposed approaches to broader applications and industries. **Dependency on Annotated Datasets:** The paper relies on manually annotated datasets for training and evaluation. Dependency on such datasets, especially in the context of sentiment analysis, may pose challenges when applying the proposed models to different domains or when dealing with evolving language trends and expressions.
Omran et al. (2023)	Product Reviews, Consumer Feedback	MSA, BDs	Noise reduce, Data augmentation, Normalizing, Removing, Tokenizing	LSTM	Accuracy: MSA: 97.01%, BDs: 96.72% F1-score: MSA: 97.69%, BDs: 97.93%	**Limited Annotated Data and Dialect Coverage:** The paper acknowledges the scarcity of annotated data in Arabic, particularly for dialects like Bahraini and Mauritanian. This limitation hinders the development and evaluation of sentiment analysis models, and the lack of coverage for specific dialects further restricts the generalizability of the findings. **Insufficient Transfer Learning and Multilingual Models:** The paper points out a deficiency in multilingual deep learning models and insufficient exploration of transfer learning for Arabic sentiment analysis, particularly for dialects.
Başarslan and Kayaalp (2023)	Entertainment	IMDB movie reviews	Data cleaning	MBi-GRUMCONV	Accuracy: 95.34%	**Unaddressed Model Drawbacks:** The paper overlooks mentioning potential limitations or drawbacks associated with the MBi-GRUMCONV model, leaving gaps in understanding its weaknesses or areas for improvement. **Limited Generalizability Discussion:** Lack of discussion on the generalizability of the model to other datasets or domains raises concerns about its adaptability beyond the specific context of IMDB movie reviews.

certain groups or sentiments. Mao et al. (2023) highlighted that pre-trained language models exhibit bias in sentiment analysis and emotion detection tasks, particularly regarding the number of label classes and the selection of emotional label words. Fairness is particularly crucial in sentiment analysis, as biased models can perpetuate stereotypes or misrepresent the sentiments of specific demographic groups. For example, if a sentiment analysis tool consistently misinterprets reviews from a particular cultural background as unfavorable due to cultural differences in expressions, it demonstrates bias. The difficulty lies in recognizing and alleviating these biases to uphold the integrity and fairness of sentiment analysis, ensuring equal representation and precise sentiment classification across all groups.

Future research in sentiment analysis should focus on developing debiasing techniques like Reweighting, Counterfactual Data Generation, Adversarial Training, etc., for sentiment models and promoting fairness in treating diverse demographic groups (Sun et al., 2023). Methods to detect and mitigate biases within sentiment models, especially pre-trained language models, should be explored. Additionally, researchers should work on creating diverse and balanced sentiment datasets that represent various cultural and demographic backgrounds. This will lead to more equitable and accurate sentiment analysis results that align with the diversity of sentiment expressions in the real world.

8.5. Explainability

Explainability in sentiment analysis pertains to the capacity to clarify the decision-making process of sentiment models and make their predictions transparent and understandable. This poses a significant challenge in sentiment analysis, particularly with advanced models like deep learning techniques, as they are often regarded as "black boxes", making it challenging to understand the rationale behind the assignment of specific sentiments (Choudhary et al., 2022). Lack of explainability can lead to skepticism and mistrust in sentiment analysis outcomes. For instance, if a sentiment model classifies a product review as unfavorable without providing clear reasons, it hinders the user's ability to understand and trust the results, impacting its reliability and utility. Achieving explainability is essential to enhance user confidence and ensure the credibility of sentiment analysis results.

Work should be done on developing interpretable deep learning models that provide insights into their decision processes. Techniques like attention mechanisms (Feng et al., 2022) and model visualization can shed light on why specific sentiments are assigned. Creating transparent model-agnostic, Explainable Artificial Intelligence (XAI) techniques with explanations (Pavitha et al., 2022b) and establishing guidelines or standards for explainable sentiment analysis can improve user trust and facilitate broader adoption of sentiment analysis in real-world applications. XAI techniques include a range of methods,

Table 16
Results analysis of state-of-the-art research articles on LLM-based sentimental analysis.

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Kojima et al. (2022)	Education, E-commerce	MultiArith	Data cleaning, Reasoning extraction, Features extraction	InstructGPT	Accuracy: MultiArith:78.7%	**Narrow Reasoning Focus:** The CRASS paper's limitation lies in its narrow focus on counterfactual reasoning, restricting the evaluation to a specific type of cognitive ability without addressing broader aspects of reasoning. **Risk of Overfitting:** The framing of tasks as counterfactual questions may pose a risk of overfitting, especially if models are extensively pre-trained on questioned counterfactual conditionals (QCCs), potentially leading to unwanted artifacts in encoding and model performance.
Huang et al. (2022a)	Product Reviews, Consumer Feedback	GSM8K, ANLI-A3, OpenBookQA	Removal, Features extraction	LMSI	Accuracy: GSM8K: 78.7%, ANLI-A3: 67.9%, OpenBookQA: 94.4%	**Unacknowledged Ethical Concerns:** The paper overlooks a discussion on potential ethical implications or risks associated with self-improvement techniques for large language models, leaving unexplored considerations related to responsible AI development. **Comparative Analysis:** The absence of a comparison with other methods for enhancing reasoning abilities in language models limits the understanding of the relative effectiveness and applicability of the self-improvement approach.
Frohberg and Binder (2021)	Social Media,E-commerce	CRASS	Features extraction	MPNet, GPT-3	Accuracy: GPT-3: 31.85%, MPNet:35.11%	Narrow Reasoning Focus: The CRASS paper is limited by its exclusive emphasis on a specific form of reasoning, namely counterfactual reasoning. This narrow focus may restrict the broader applicability of the benchmark to other types of reasoning abilities. **Potential for Overfitting:** Framing the task as a counterfactual question might introduce challenges related to overfitting. If models are extensively pre-trained on a multitude of questionized counterfactual conditionals (QCCs), there is a risk that the models may become overly specialized and fail to generalize effectively to a wider range of reasoning tasks.
Hong et al. (2023)	Financial Markets, E-commerce	3D-LLMs	Features extraction	ScanQA	BLEU-1: 9%	Rendering Process Overhead: The 3D-LLM paper's reliance on 2D multi-view images for the 3D feature extractor necessitates an additional rendering process for training. This introduces an overhead in terms of computational resources and time. **Lack of Computational Analysis:** The paper lacks a detailed analysis of the computational cost or efficiency associated with the proposed 3D-LLMs. The absence of such insights hinders a comprehensive understanding of the practical feasibility and scalability of the approach.
Fatouros et al. (2023)	Financial Markets	ForexLive, FXstreet	Tokenization, Features extraction	FinBERT, GPT	Accuracy: 56.1%, Precision: 56.0%, Recall: 56.2%, F1-Score: 55.6%, S-MAE: 54.0%	**Limited Temporal Analysis:** The paper lacks exploration of the temporal relationship between sentiment scores derived from ChatGPT and subsequent market movements. This limitation restricts insights into the time-dependent dynamics of financial sentiment. **Short Time Frame Bias:** The study's reliance on a specific time frame may introduce bias, and the models' performance might not generalize well to different periods or diverse market conditions. This raises concerns about the robustness and applicability of the findings across various temporal scenarios.
Okey et al. (2023)	Healthcare	Social Media	Remove noise, Features extraction, Tokenization	InstructGPT, BERT	Positive 43.8%, Neutral 36.3%, Negative 19.9%	**Data Source Limitation:** The study relies exclusively on Twitter data for sentiment analysis, potentially limiting the diversity of opinions and perspectives to those expressed on this platform. This may not capture a comprehensive view of public sentiment on ChatGPT and cybersecurity. **Sentiment Analysis Model Dependency:** The paper employs specific sentiment analysis models (VADER and roBERTa) without exploring potential variations in results due to the choice of models. The generalizability of findings may be influenced by the performance characteristics and biases inherent in these particular models.
Austin et al. (2021)	Education	MBPP	Tokenization, Removal, Feature Extraction	MathQA-Python	Accuracy: 83.8%	**Benchmark Limitation:** The benchmark programs utilized are criticized for their simplicity and brevity, potentially undermining the ability to assess the models' performance on more extensive and intricate program synthesis tasks. **Limited Predictive Ability:** The paper acknowledges a general limitation in the models' capability to predict program output for a given input, even when employing state-of-the-art models, indicating a broader challenge in achieving accurate predictions.

(*continued on next page*)

Table 16 (*continued*).

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Bansal et al. (2024)	Healthcare	GSM-8K	Tokenization, Feature extraction	CALM	Accuracy: 84.3%	**Restricted Modification:** Imposes a constraint on not modifying the weights of either the anchor or augmenting models during composition. This restriction might limit the adaptability of CALM in scenarios where more flexible modifications could enhance performance. **Small Data Dependency:** CALM assumes access to a small amount of data representing the combined skills of the given models. Dependency on limited data might be a limitation, especially in cases where obtaining representative combined skills data is challenging or not feasible.

Table 17
Results analysis of state-of-the-art research articles on Pre-trained-based sentimental analysis.

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Tarcar et al. (2019)	Healthcare	2300 sample notes from the Medical Transcripts Samples site and 100 FAQ sections from the EMA site.	Tokenization, Features extraction, Features selection	NER	F1 Score: 73.4%	**Dataset Size Uncertainty:** The absence of information about the size of the dataset used for training the NER model raises concerns about the generalizability of the results, as the effectiveness of the model could be influenced by the dataset's scale. **Limited Evaluation Metrics Diversity:** The paper's exclusive reliance on the F1 score as the sole evaluation metric restricts the comprehensive assessment of the NER model's performance, potentially overlooking aspects captured by other relevant metrics.
Ji et al. (2021)	Healthcare	Reddit, Twitter	Tokenization, Removing, Data cleaning	MentalBERT, MentalRoBERTa	MentalRoBERTa: F1 Score: 94.62%, MentalBERT: F1 Score: 81.76%	**Limited Exploration of Pretraining Techniques:** The paper focuses on standard pretraining protocols for MentalBERT and MentalRoBERTa, lacking exploration of novel pretraining techniques specific to mental healthcare, potentially limiting the models' adaptability to the nuances of mental health-related text. **Resource-Intensive Training:** The computational resources required for training the language models are substantial, using four Nvidia Tesla v100 GPUs for eight days. This resource demand may hinder broader accessibility and adoption of the pre-trained models in resource-constrained environments.
Araci (2019)	Financial Markets	TRC2-financial	Tokenization, Normalization, Padding	FinBERT	Accuracy: 97.0% F1 Score: 95.0%	**Limited Exploration of Model Variants:** The paper primarily focuses on introducing FinBERT based on BERT and evaluates its performance against state-of-the-art methods. However, it may lack a comprehensive exploration of alternative model architectures or variations beyond BERT for financial sentiment analysis. **Evaluation on Limited Datasets:** The paper mentions the evaluation of FinBERT on two financial sentiment analysis datasets. However, the limitations could arise from potential biases or specific characteristics of these datasets, and the generalizability of FinBERT to a broader range of financial text data remains to be thoroughly investigated.
Lamsal et al. (2023)	Social Media	Twitter	Embedding, Feature Scaling	Transformers	F1 Score: 95.0%	**Limited Comparative Analysis:** While the paper introduces Crisis Transformers and evaluates its performance against existing models and baselines, it might lack a comprehensive comparative analysis with other state-of-the-art models in related domains. A broader evaluation could provide a more nuanced understanding of the model's strengths and weaknesses. **Dependency on Social Media Data:** The reliance on crisis-related social media texts, particularly from platforms like Twitter, might introduce a bias in the training and evaluation of CrisisTransformers.

(*continued on next page*)

Table 17 (*continued*).

Ref.	Application	Dataset	Pre-Processing method	Model	Results	Limitations
Liu et al. (2023)	Education, Healthcare	SR1, SR2	Feature Scaling, Tokenization	BERT	Accuracy: SR1-BERT: 91.2%, SR2-BERT: 90.4%	**Limited Analysis of Model Generalization:** The paper may have a limitation in thoroughly analyzing the generalization of SciEdBERT to a broader set of science education tasks. While the study focuses on specific downstream tasks related to scientific argumentation, it might benefit from investigating the model's performance on a more diverse range of assessment types to assess its robustness and applicability. **Potential Bias in Training Data:** The reliance on a large dataset of students' written responses raises concerns about potential biases in the training data, such as regional or institutional variations. The paper could address how the model's performance might be influenced by the characteristics of the student population contributing to the dataset, ensuring transparency about the potential limitations in terms of representativeness.
Kumar et al. (2023)	Social Media	CelebA	Feature extraction, Removal, Tokenization	GAN	Accuracy: 81.25%	**Limited Discussion on Model Robustness:** The paper lacks a detailed discussion on the robustness limitations of the GAN-based deepfake detection model. It does not thoroughly explore potential vulnerabilities or failure cases that could affect the model's performance in real-world scenarios beyond the training conditions. **Omission of Dataset Limitations:** The paper overlooks a critical aspect by not addressing the limitations of the celebA dataset used for training and evaluation. This dataset's representational bias, focusing on celebrity faces, might impact the model's generalizability to diverse real-world situations, limiting its applicability beyond specific contexts.
Jain (2021b)	Customer Support	Customer SupportChat	Tokenization, Lowercasing, Removal of Stop Words	RoBERTa	F1 Score: 65.0%	**Limited Generalizability:** The paper focuses on sentiment analysis in the specific context of customer-agent chat data. The proposed model's applicability may be limited to this domain, and its performance might not generalize well to other types of sentiment analysis tasks or datasets. **Dependency on Weak Supervision Techniques:**The paper heavily relies on weak supervision techniques, such as domain-specific lexicon-based rules and weak classifiers, for training the sentiment analysis model. The effectiveness of the model is intricately tied to the quality and relevance of these weak labels, which may introduce noise and impact the model's robustness in real-world scenarios.

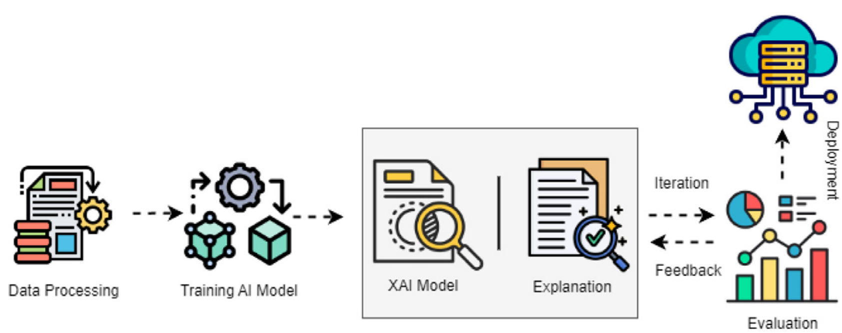

Fig. 13. Working procedure of Explainable Artificial Intelligence model.

such as rule-based systems, feature importance analysis, and model interpretability approaches like Local Interpretable Model-Agnostic Explanations (LIME) and Shapley values (Minh et al., 2022). Fig. 13 depicts the operational process of explainable artificial intelligence. These approaches aim to bridge the gap between complex models and user understanding, ensuring that sentiment analysis results are trustworthy and interpretable.

8.6. Privacy concerns

Privacy concerns in sentiment analysis revolve around collecting and using personal data, as sentiment analysis often relies on text data from various sources. Analyzing this data can pose privacy risks when users' information, emotions, or opinions are processed without their consent or knowledge (Lee et al., 2022). This challenge is significant in sentiment analysis because the ethical and legal implications of handling private data must be addressed. For instance, analyzing user comments on social media for sentiment may inadvertently expose sensitive personal information, leading to privacy breaches and potential legal repercussions.

Developing advanced privacy-preserving techniques, like Federated Learning (FL), can help protect user data while enabling sentiment analysis to overcome privacy concerns in sentiment analysis. FL enables AI model training using decentralized user data without sharing the raw data. During this procedure, the local model of each user is transmitted to the central server for training, as depicted in Fig. 14; instead of sending the user's data directly, the approach involves transmitting the

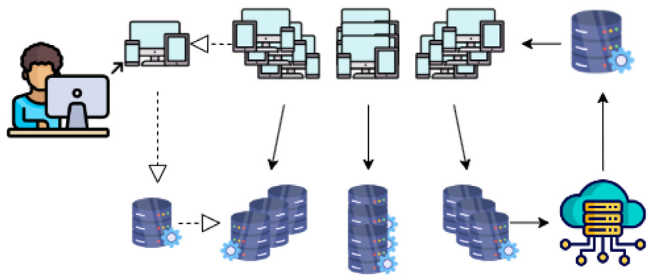

Fig. 14. Working procedure of Federated Learning.

user's model. While researchers are beginning to explore FL in Sentiment Analysis, it is still in its early stages (Bansal et al., 2022), Nagy et al. (2023), Sakhare and Shaik (2023). A significant challenge is ensuring efficient communication and data privacy while synchronizing model updates across distributed devices. Future research should focus on overcoming these issues related to scalability, convergence speed, privacy in distributed training, and enhancing sentiment analysis efficiency. Differentiated privacy methods, data noise addition, robust authentication, and user-friendly privacy controls are critical for safeguarding user privacy. Additionally, establishing robust data anonymization and consent mechanisms is crucial to safeguard personal information during sentiment analysis. Researchers should work on ethical guidelines and frameworks for responsible sentiment analysis, ensuring compliance with privacy regulations and minimizing the risk of data exposure.

8.7. Multimodal sentiment analysis

Multimodal Sentiment Analysis is a complex field that involves analyzing sentiment in data that includes multiple modes, such as text, images, and audio. The major challenge in sentiment analysis is the fusion and interpretation of information from these diverse sources, as each mode provides distinct cues about sentiment (Gandhi et al., 2023), (Das and Singh, 2023). Practical difficulties arise when analyzing user-generated content on a social platform incorporating text comments, images, and voice recordings. The challenge lies in integrating these modalities effectively to obtain a holistic understanding of sentiment, as different modes may convey contrasting or complementary emotional cues. Addressing this challenge is crucial for comprehensive sentiment analysis across various data sources.

To overcome the challenges of Multimodal Sentiment Analysis, the focus should be on developing advanced techniques like Multimodal Neural Networks (MMNs) (Choksi et al., 2022) or Multimodal Transformers (Yu et al., 2022) for adeptly integrating and deciphering data from diverse modalities, including text, images, and audio. The challenge lies in seamlessly integrating these diverse sources to comprehensively understand sentiment, especially in social media content with multiple modes. Researchers should explore innovative approaches to address the fusion of contrasting or complementary emotional cues from different modalities, enhancing the capabilities of sentiment analysis across various data sources.

8.8. Long-Term Context Understanding

Long-Term Context Understanding in sentiment analysis is a vital aspect that involves the capability to take into account extended or historical context when assessing the sentiment of a text or conversation. This presents a substantial challenge because sentiments can change and evolve, and interpreting the sentiment of a particular statement may necessitate an awareness of preceding interactions or events (Huang et al., 2020). For instance, a user's response may refer to prior messages or issues in social media discussions or customer service interactions. Without access to this long-term context, sentiment analysis risks inaccuracies and misinterpretations.

Long-term context Understanding for sentiment analysis should focus on developing advanced algorithms capable of effectively capturing and incorporating historical context, such as Dynamic Memory Networks, Transformer-based Models, Hierarchical Models and Reinforcement Learning-based Models. Though research is going on integrating these methods in sentiment analysis (Wen et al., 2023), (Kokab et al., 2022), (Yu et al., 2022), (Cao et al., 2023), they should be more developed to identify and analyze the changing sentiment trajectories and understand how past interactions or events influence current sentiment expressions. Additionally, research can explore techniques considering the context's duration and significance, leading to more accurate and nuanced sentiment interpretations, particularly in complex conversations or long-term interactions.

8.9. Resource constraints

Resource Constraints are a notable challenge in sentiment analysis, as they encompass limitations in computational power, data availability, and human resources. Performing sentiment analysis often demands substantial computational resources, particularly when employing deep learning models and processing large datasets, making it challenging for individuals and organizations with limited computing capabilities. Additionally, acquiring labeled data for training sentiment models can be resource-intensive, and the shortage of such data can hinder model development. Moreover, manual labeling by human annotators is time-consuming and expensive, and resource constraints can limit the extent and accuracy of the labeling process. These challenges can result in less accurate sentiment analysis outcomes and hinder the deployment of sentiment analysis solutions, particularly for small businesses or projects with constrained resources.

Therefore, research should be done on developing lightweight sentiment analysis models that are efficient in terms of computational power and memory usage. Additionally, creating semi-supervised or unsupervised learning approaches can reduce the dependency on large labeled datasets, making sentiment analysis more accessible to a broader range of users. Leveraging active learning techniques can optimize the manual labeling process, maximizing the use of limited human resources for data annotation. Furthermore, exploring transfer learning and domain adaptation methods can help adapt sentiment models to low-resource languages and domains. Reconfigurable computing can also enhance sentiment analysis by optimizing resource utilization, thereby addressing resource constraints. These efforts will ensure that sentiment analysis remains viable for a wider audience, even when faced with resource constraints.

9. Conclusions

Sentiment analysis determines and categorizes a text's emotional tone or attitude, such as positive, negative, or neutral. It is crucial for understanding public opinion, customer feedback, and market trends, enabling businesses and organizations to make data-driven decisions and enhance user experiences. However, recent studies in this extensive research area are insufficient, prompting this investigation to undertake a systematic review to provide in-depth insights into sentiment analysis and its recent advancements. It covered many sentiment analysis applications, commonly used data preprocessing methods, and experimental datasets while examining the strengths and limitations of prevalent ML and DL algorithms. Additionally, it offered a thorough exploration and analysis of results from recent state-of-the-art research articles. The paper also addressed the challenges in sentiment analysis and outlined future research directions to overcome these hurdles. This comprehensive review serves as a valuable interdisciplinary resource and provides essential insights for advancing the field of sentiment analysis, ultimately benefiting investors and researchers by promoting innovation and well-informed decision-making.

CRediT authorship contribution statement

Jamin Rahman Jim: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Md Apon Riaz Talukder:** Data curation, Formal analysis, Methodology, Visualization, Writing – original draft. **Partha Malakar:** Data curation, Formal analysis, Methodology, Writing – original draft. **Md Mohsin Kabir:** Formal analysis, Methodology, Supervision, Validation, Writing – review & editing. **Kamruddin Nur:** Investigation, Supervision, Writing – review & editing. **M.F. Mridha:** Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. M. F. Mridha reports was provided by American International University Bangladesh. Dr. M. F. Mridha reports a relationship with American International University Bangladesh that includes: employment.

Acknowledgment

The authors would like to thank the Advanced Machine Intelligence Research Lab - AMIR Lab for Supervision and Resources.

References

Abdullah, T., Ahmet, A., 2022. Deep learning in sentiment analysis: Recent architectures. ACM Comput. Surv. 55 (8), 1–37.
Ahangari, M., Sebti, A., 2023. A hybrid approach to sentiment analysis of Iranian stock market user's opinions. Int. J. Eng. 36 (3), 573–584.
Ahmed, A.A.A., Agarwal, S., Kurniawan, I.G.A., Anantadjaya, S.P., Krishnan, C., 2022. Business boosting through sentiment analysis using artificial intelligence approach. Int. J. Syst. Assur. Eng. Manag. 13 (Suppl 1), 699–709.
Ahmed, C., ElKorany, A., ElSayed, E., 2023. Prediction of customer's perception in social networks by integrating sentiment analysis and machine learning. J. Intell. Inf. Syst. 60 (3), 829–851.
Al-Mashhadany, A.K., Sadiq, A.T., Ali, S.M., Ahmed, A.A., 2022. Healthcare assessment for beauty centers using hybrid sentiment analysis. Indonesian J. Electr. Eng. Comput. Sci. 28 (2), 890–897.
Alamoudi, E.S., Alghamdi, N.S., 2021. Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. J. Decis. Syst. 30 (2–3), 259–281.
Alaparthi, S., Mishra, M., 2020. Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. arXiv preprint arXiv:2007.01127.
Alassaf, M., Qamar, A.M., 2022. Improving sentiment analysis of Arabic tweets by one-way ANOVA. J. King Saud Univ.-Comput. Inf. Sci. 34 (6), 2849–2859.
Ali, P.J.M., Faraj, R.H., Koya, E., Ali, P.J.M., Faraj, R.H., 2014. Data normalization and standardization: A technical report. Mach. Learn. Tech. Rep. 1 (1), 1–6.
Aliwy, A.H., 2012. Tokenization as preprocessing for Arabic tagging system. Int. J. Inf. Educ. Technol. 2 (4), 348.
AlSulaim, S.M., Qamar, A.M., 2021. Prediction of anime series' success using sentiment analysis and deep learning. In: 2021 International Conference of Women in Data Science at Taif University. WiDSTaif, IEEE, pp. 1–6.
Araci, D., 2019. Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.
Araújo, M., Pereira, A., Benevenuto, F., 2020. A comparative study of machine translation for multilingual sentence-level sentiment analysis. Inform. Sci. 512, 1078–1102.
Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H., Shah, J., 2020. Sentiment analysis of extremism in social media from textual information. Telemat. Inform. 48, 101345.
Aslan, S., Kızıloluk, S., Sert, E., 2023. TSA-CNN-AOA: Twitter sentiment analysis using CNN optimized via arithmetic optimization algorithm. Neural Comput. Appl. 1–18.
Austin, J., Odena, A., Nye, M., Bosma, M., Michalewski, H., Dohan, D., Jiang, E., Cai, C., Terry, M., Le, Q., et al., 2021. Program synthesis with large language models. arXiv preprint arXiv:2108.07732.
Aziz, A., Chowdhury, N.K., Kabir, M.A., Chy, A.N., Siddique, M.J., 2023. MMTF-DES: A fusion of multimodal transformer models for desire, emotion, and sentiment analysis of social media data. arXiv preprint arXiv:2310.14143.
Balaji, P., Nagaraju, O., Haritha, D., 2017. Levels of sentiment analysis and its challenges: A literature review. In: 2017 International Conference on Big Data Analytics and Computational Intelligence. ICBDAC, IEEE, pp. 436–439.
Bank, D., Koenigstein, N., Giryes, R., 2023. Autoencoders. In: Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook. Springer, pp. 353–374.
Bansal, A., Kumar, N., 2022. Aspect-based sentiment analysis using attribute extraction of hospital reviews. New Gener. Comput. 1–20.
Bansal, R., Samanta, B., Dalmia, S., Gupta, N., Vashishth, S., Ganapathy, S., Bapna, A., Jain, P., Talukdar, P., 2024. LLM augmented LLMs: Expanding capabilities through composition. arXiv preprint arXiv:2401.02412.
Bansal, S., Singh, M., Bhadauria, M., Adalakha, R., 2022. Federated learning approach towards sentiment analysis. In: 2022 2nd International Conference on Technological Advancements in Computational Sciences. ICTACS, pp. 717–724. http://dx.doi.org/10.1109/ICTACS56270.2022.9987996.
Başarslan, M.S., Kayaalp, F., 2023. MBi-GRUMCONV: A novel multi bi-GRU and multi CNN-based deep learning model for social media sentiment analysis. J. Cloud Comput. 12 (1), 5.
Benrouba, F., Boudour, R., 2023. Emotional sentiment analysis of social media content for mental health safety. Soc. Netw. Anal. Min. 13 (1), 17.
Bhatia, S., Kumar, A., Khan, M.M., 2022. Role of genetic algorithm in optimization of Hindi word sense disambiguation. IEEE Access 10, 75693–75707.
Birjali, M., Kasri, M., Beni-Hssane, A., 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowl.-Based Syst. 226, 107134.
Borg, A., Boldt, M., 2020. Using VADER sentiment and SVM for predicting customer response sentiment. Expert Syst. Appl. 162, 113746.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A., 2017. Affective computing and sentiment analysis. In: A Practical Guide to Sentiment Analysis. Springer, pp. 1–10.
Cambria, E., Hazarika, D., Poria, S., Hussain, A., Subramanyam, R., 2018. Benchmarking multimodal sentiment analysis. In: Computational Linguistics and Intelligent Text Processing: 18th International Conference, CICLing 2017, Budapest, Hungary, April 17–23, 2017, Revised Selected Papers, Part II 18. Springer, pp. 166–179.
Cao, Y., Tang, Y., Du, H., Xu, F., Wei, Z., Jin, C., 2023. Heterogeneous reinforcement learning network for aspect-based sentiment classification with external knowledge. IEEE Trans. Affect. Comput..
Capuano, N., Greco, L., Ritrovato, P., Vento, M., 2021. Sentiment analysis for customer relationship management: An incremental learning approach. Appl. Intell. 51, 3339–3352.
Carvalho, I., Silva, C., Oliveira, H.G., 2023. The importance of context for sentiment analysis in dialogues. IEEE Access.
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al., 2015. Xgboost: Extreme gradient boosting. pp. 1–4, R package version 0.4-2, 1, 4.
Chen, L.-C., Lee, C.-M., Chen, M.-Y., 2020. Exploration of social media for sentiment analysis using deep learning. Soft Comput. 24, 8187–8197.
Choksi, B., Mozafari, M., Vanrullen, R., Reddy, L., 2022. Multimodal neural networks better explain multivoxel patterns in the hippocampus. Neural Netw. 154, 538–542.
Choudhary, S., Chatterjee, N., Saha, S.K., 2022. Interpretation of black box NLP models: A survey. arXiv preprint arXiv:2203.17081.
Chu, E., Roy, D., 2017. Audio-visual sentiment analysis for learning emotional arcs in movies. In: 2017 IEEE International Conference on Data Mining. ICDM, IEEE, pp. 829–834.
Dake, D.K., Gyimah, E., 2023. Using sentiment analysis to evaluate qualitative students' responses. Educ. Inf. Technol. 28 (4), 4629–4647.
Das, R., Singh, T.D., 2023. Multimodal sentiment analysis: A survey of methods, trends and challenges. ACM Comput. Surv..
Deepa, N., Priya, J.S., Devi, T., 2023. Sentimental analysis recognition in customer review using novel-CNN. In: 2023 International Conference on Computer Communication and Informatics. ICCCI, IEEE, pp. 1–4.
Demircan, M., Seller, A., Abut, F., Akay, M.F., 2021. Developing turkish sentiment analysis models using machine learning and e-commerce data. Int. J. Cogn. Comput. Eng. 2, 202–207.
Devika, M.D., Sunitha, C., Ganesh, A., 2016. Sentiment analysis: A comparative study on different approaches. Procedia Comput. Sci. 87, 44–49.
Diekson, Z.A., Prakoso, M.R.B., Putra, M.S.Q., Syaputra, M.S.A.F., Achmad, S., Sutoyo, R., 2023. Sentiment analysis for customer review: Case study of Traveloka. Procedia Comput. Sci. 216, 682–690.
El-Ansari, A., Beni-Hssane, A., 2023. Sentiment analysis for personalized chatbots in e-commerce applications. Wirel. Pers. Commun. 129 (3), 1623–1644.
Fang, Z., Lu, J., Liu, F., Zhang, G., 2022. Semi-supervised heterogeneous domain adaptation: Theory and algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 45 (1), 1087–1105.
Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., Kyriazis, D., 2023. Transforming sentiment analysis in the financial domain with ChatGPT. Mach. Learn. Appl. 14, 100508.
Feng, A., Zhang, X., Song, X., 2022. Unrestricted attention may not be all you need–masked attention mechanism focuses better on relevant parts in aspect-based sentiment analysis. IEEE Access 10, 8518–8528.
Floridi, L., Chiriatti, M., 2020. GPT-3: Its nature, scope, limits, and consequences. Minds Mach. 30, 681–694.
Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learn. 29, 131–163.

Frohberg, J., Binder, F., 2021. Crass: A novel data set and benchmark to test counterfactual reasoning of large language models. arXiv preprint arXiv:2112.11941.
Gamberger, D., Lavrac, N., Dzeroski, S., 2000. Noise detection and elimination in data preprocessing: Experiments in medical domains. Appl. Artif. Intell. 14 (2), 205–223.
Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., Hussain, A., 2023. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Inf. Fusion 91, 424–444.
Ganesan, S., Somasiri, N., Colombage, C., 2023. Deep learning approaches for accurate sentiment analysis of online consumer feedback. In: 2023 International Conference on Computer Communication and Informatics. ICCCI, IEEE, pp. 1–5.
Ghosal, S., Jain, A., 2023. Weighted aspect based sentiment analysis using extended OWA operators and Word2Vec for tourism. Multimedia Tools Appl. 82 (12), 18353–18380.
Ghosh, R., Ravi, K., Ravi, V., 2016. A novel deep learning architecture for sentiment classification. In: 2016 3rd International Conference on Recent Advances in Information Technology. RAIT, IEEE, pp. 511–516.
Ghosh, A., Umer, S., Khan, M.K., Rout, R.K., Dhara, B.C., 2023. Smart sentiment analysis system for pain detection using cutting edge techniques in a smart healthcare framework. Cluster Comput. 26 (1), 119–135.
Grover, V., 2022. Exploiting emojis in sentiment analysis: A survey. J. Inst. Eng. (India): Ser. B 103 (1), 259–272.
Guo, Y., et al., 2022. Financial market sentiment prediction technology and application based on deep learning model. Comput. Intell. Neurosci. 2022.
Gupta, S., Arora, B., 2022. Stemming techniques on english language and devanagari script: A review. In: Recent Innovations in Computing: Proceedings of ICRIC 2021, vol. 1, Springer, pp. 541–550.
Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabduallah, B.I., Siddiqui, T., Albahlal, B.M., Alajlan, S.A., Al-Subaie, M., 2023. DDPM: A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms. Diagnostics 13 (6), 1093.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3 (Mar), 1157–1182.
Guyon, I., Elisseeff, A., 2006. An introduction to feature extraction. In: Feature Extraction: Foundations and Applications. Springer, pp. 1–25.
Halawani, H.T., Mashraqi, A.M., Badr, S.K., Alkhalaf, S., 2023. Automated sentiment analysis in social media using Harris Hawks optimisation and deep learning techniques. Alex. Eng. J. 80, 433–443.
Hayawi, K., Shahriar, S., Serhani, M.A., Taleb, I., Mathew, S.S., 2022. ANTi-Vax: A novel Twitter dataset for COVID-19 vaccine misinformation detection. Public Health 203, 23–30.
Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intell. Syst. Appl. 13 (4), 18–28.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hong, Y., Zhen, H., Chen, P., Zheng, S., Du, Y., Chen, Z., Gan, C., 2023. 3D-llm: Injecting the 3d world into large language models. arXiv preprint arXiv:2307.12981.
Hu, A., Flaxman, S., 2018. Multimodal sentiment analysis to explore the structure of emotions. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 350–358.
Huang, H., Asemi, A., Mustafa, M.B., 2023. Sentiment analysis in e-commerce platforms: A review of current techniques and future directions. IEEE Access.
Huang, J., Gu, S.S., Hou, L., Wu, Y., Wang, X., Yu, H., Han, J., 2022a. Large language models can self-improve. arXiv preprint arXiv:2210.11610.
Huang, W., Lin, M., Wang, Y., 2022b. Sentiment analysis of Chinese e-commerce product reviews using ERNIE word embedding and attention mechanism. Appl. Sci. 12 (14), 7182.
Huang, F., Yuan, C., Bi, Y., Lu, J., 2020. Exploiting long-term dependency for topic sentiment analysis. IEEE Access 8, 221963–221974.
Huang, F., Zhang, X., Zhao, Z., Xu, J., Li, Z., 2019. Image–text sentiment analysis via deep multimodal attentive fusion. Knowl.-Based Syst. 167, 26–37.
Hussein, D.M.E.-D.M., 2018. A survey on sentiment analysis challenges. J. King Saud Univ., Eng. Sci. 30 (4), 330–338.
Jain, N., 2021a. Customer sentiment analysis using weak supervision for customer-agent chat. arXiv preprint arXiv:2111.14282.
Jain, N., 2021b. Customer sentiment analysis using weak supervision for customer-agent chat. arXiv preprint arXiv:2111.14282.
Jain, R., Kumar, A., Nayyar, A., Dewan, K., Garg, R., Raman, S., Ganguly, S., 2023. Explaining sentiment analysis results on social media texts through visualization. Multimedia Tools Appl. 1–17.
Jaynes, E.T., 1982. On the rationale of maximum-entropy methods. Proc. IEEE 70 (9), 939–952.
Jbene, M., Tigani, S., Chehri, A., Saadane, R., et al., 2022. User sentiment analysis in conversational systems based on augmentation and attention-based BiLSTM. Procedia Comput. Sci. 207, 4106–4112.
Jha, B.K., Sivasankari, G., Venugopal, K., 2021. Sentiment analysis for e-commerce products using natural language processing. Ann. Roman. Soc. Cell Biol. 166–175.
Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., Cambria, E., 2021. Mentalbert: Publicly available pretrained language models for mental healthcare. arXiv preprint arXiv:2110.15621.
Joel, L.O., Doorsamy, W., Paul, B.S., 2022. A review of missing data handling techniques for machine learning. Int. J. Innov. Technol. Interdiscip. Sci. 5 (3), 971–1005.
Kabir, M., Mahfuz, O.B., Raiyan, S.R., Mahmud, H., Hasan, M.K., 2023. BanglaBook: A large-scale bangla dataset for sentiment analysis from book reviews. arXiv preprint arXiv:2305.06595.
Karn, A.L., Karna, R.K., Kondamudi, B.R., Bagale, G., Pustokhin, D.A., Pustokhina, I.V., Sengan, S., 2023. Customer centric hybrid recommendation system for e-commerce applications by integrating hybrid sentiment analysis. Electron. Commerce Res. 23 (1), 279–314.
Kaur, G., Sharma, A., 2023. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J. Big Data 10 (1), 5.
Keele, S., et al., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report, Ver. 2.3 Ebse Technical Report. Ebse.
Kheiri, K., Karimi, H., 2023. Sentimentgpt: Exploiting gpt for advanced sentiment analysis and its departure from current machine learning. arXiv preprint arXiv:2307.10234.
Kitchenham, B., 2004. Procedures for Performing Systematic Reviews, vol. 33, (no. 2004), Keele, UK, Keele University, pp. 1–26.
Kojima, T., Gu, S.S., Reid, M., Matsuo, Y., Iwasawa, Y., 2022. Large language models are zero-shot reasoners. In: Advances in Neural Information Processing Systems, vol. 35, pp. 22199–22213.
Kokab, S.T., Asghar, S., Naz, S., 2022. Transformer-based deep learning models for the sentiment analysis of social media data. Array 14, 100157.
Kora, R., Mohammed, A., 2023. An enhanced approach for sentiment analysis based on meta-ensemble deep learning. Soc. Netw. Anal. Min. 13 (1), 38.
Kuchaiev, O., Ginsburg, B., 2017. Training deep autoencoders for collaborative filtering. arXiv preprint arXiv:1708.01715.
Kumar, S., De, K., Roy, P.P., 2020. Movie recommendation system using sentiment analysis from microblogging data. IEEE Trans. Comput. Soc. Syst. 7 (4), 915–923.
Kumar, M., Sharma, H.K., et al., 2023. A GAN-based model of deepfake detection in social media. Procedia Comput. Sci. 218, 2153–2162.
Lamsal, R., Read, M.R., Karunasekera, S., 2023. CrisisTransformers: Pre-trained language models and sentence encoders for crisis-related social media texts. arXiv preprint arXiv:2309.05494.
Lee, C.B., Io, H.N., Tang, H., 2022. Sentiments and perceptions after a privacy breach incident. Cogen. Bus. Manag. 9 (1), 2050018.
Lei, S., Dong, G., Wang, X., Wang, K., Wang, S., 2023. Instructerc: Reforming emotion recognition in conversation with a retrieval multi-task llms framework. arXiv preprint arXiv:2309.11911.
Li, Z., Xie, H., Cheng, G., Li, Q., 2021. Word-level emotion distribution with two schemas for short text emotion classification. Knowl.-Based Syst. 227, 107163.
Li, X., Zhang, J., Du, Y., Zhu, J., Fan, Y., Chen, X., 2023. A novel deep learning-based sentiment analysis method enhanced with Emojis in microblog social networks. Enterprise Inf. Syst. 17 (5), 2037160.
Ligthart, A., Catal, C., Tekinerdogan, B., 2021. Systematic reviews in sentiment analysis: A tertiary study. Artif. Intell. Rev. 1–57.
Liu, B., 2022. Sentiment Analysis and Opinion Mining. Springer Nature.
Liu, Z., He, X., Liu, L., Liu, T., Zhai, X., 2023. Context matters: A strategy to pre-train language model for science education. arXiv preprint arXiv:2301.12031.
Lu, Q., Sun, X., Long, Y., Gao, Z., Feng, J., Sun, T., 2023. Sentiment analysis: Comprehensive reviews, recent advances, and open challenges. IEEE Trans. Neural Netw. Learn. Syst..
Ma, J., Rong, L., Zhang, Y., Tiwari, P., 2023. Moving from narrative to interactive multi-modal sentiment analysis: A survey. ACM Trans. Asian Low-Resource Lang. Inf. Process..
Maas, A., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C., 2011. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. pp. 142–150.
Mabokela, K.R., Celik, T., Raborife, M., 2022. Multilingual sentiment analysis for under-resourced languages: A systematic review of the landscape. IEEE Access.
Maharana, K., Mondal, S., Nemade, B., 2022. A review: Data pre-processing and data augmentation techniques. Global Transit. Proc. 3 (1), 91–99.
Mamta, Ekbal, A., Bhattacharyya, P., 2022. Exploring multi-lingual, multi-task, and adversarial learning for low-resource sentiment analysis. Trans. Asian Low-Resour. Lang. Inf. Process. 21 (5), 1–19.
Mandvikar, S., 2023. Augmenting intelligent document processing (IDP) workflows with contemporary Large Language Models (LLMs). Int. J. Comput. Trends Technol. 71 (10), 80–91.
Mao, R., Liu, Q., He, K., Li, W., Cambria, E., 2023. The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection. IEEE Trans. Affect. Comput. 14 (3), 1743–1753. http://dx.doi.org/10.1109/TAFFC.2022.3204972.
Medsker, L.R., Jain, L., 2001. Recurrent neural networks. Des. Appl. 5 (64–67), 2.
Mehmood, S., Ahmad, I., Khan, M.A., Khan, F., Whangbo, T., 2022. Sentiment analysis in social media for competitive environment using content analysis. Comput. Mater. Continua 71 (3).
Memiş, E., Akarkamçı, H., Yeniad, M., Rahebi, J., Lopez-Guede, J.M., 2024. Comparative study for sentiment analysis of financial tweets with deep learning methods. Appl. Sci. 14 (2), 588.

Memon, S.A., Carley, K.M., 2020. Characterizing COVID-19 misinformation communities using a novel twitter dataset. arXiv preprint arXiv:2008.00791.
Minh, D., Wang, H.X., Li, Y.F., Nguyen, T.N., 2022. Explainable artificial intelligence: A comprehensive review. Artif. Intell. Rev. 1–66.
Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T., Trajanov, D., 2020a. Evaluation of sentiment analysis in finance: From lexicons to transformers. IEEE Access 8, 131662–131682.
Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T., Trajanov, D., 2020b. Evaluation of sentiment analysis in finance: From lexicons to transformers. IEEE Access 8, 131662–131682.
Model, L.L.-S.P.-t., 2023. Interactive storyboarding system leveraging large-scale pre-trained model.
Modi, A., Shah, K., Shah, S., Patel, S., Shah, M., 2022. Sentiment analysis of Twitter feeds using flask environment: A superior application of data analysis. Ann. Data Sci. 1–22.
Mohammed, A., Kora, R., 2019. Deep learning approaches for Arabic sentiment analysis. Soc. Netw. Anal. Min. 9, 1–12.
Nagy, B., Hegedűs, I., Sándor, N., Egedi, B., Mehmood, H., Saravanan, K., Lóki, G., Kiss, Á., 2023. Privacy-preserving federated learning and its application to natural language processing. Knowl.-Based Syst. 268, 110475.
Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., et al., 2021. Efficient large-scale language model training on gpu clusters using Megatron-Lm. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. pp. 1–15.
Nguyen, T.Q., Salazar, J., 2019. Transformers without tears: Improving the normalization of self-attention. arXiv preprint arXiv:1910.05895.
Okey, O.D., Udo, E.U., Rosa, R.L., Rodríguez, D.Z., Kleinschmidt, J.H., 2023. Investigating ChatGPT and cybersecurity: A perspective on topic modeling and sentiment analysis. Comput. Secur. 135, 103476.
Omran, T.M., Sharef, B.T., Grosan, C., Li, Y., 2023. Transfer learning and sentiment analysis of Bahraini dialects sequential text data using multilingual deep learning approach. Data Knowl. Eng. 143, 102106.
O'Shea, K., Nash, R., 2015. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.
Ouyang, X., Zhou, P., Li, C.H., Liu, L., 2015. Sentiment analysis using convolutional neural network. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. IEEE, pp. 2359–2364.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al., 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906.
Paramesha, K., Gururaj, H., Nayyar, A., Ravishankar, K., 2023. Sentiment analysis on cross-domain textual data using classical and deep learning approaches. Multimedia Tools Appl. 1–24.
Patrick, M.K., Adekoya, A.F., Mighty, A.A., Edward, B.Y., 2022. Capsule networks–A survey. J. King Saud Univ.-Comput. Inf. Sci. 34 (1), 1295–1310.
Patro, S., Sahu, K.K., 2015. Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462.
Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., Shashidhar, R., 2022a. Movie recommendation and sentiment analysis using machine learning. Global Transit. Proc. 3 (1), 279–284.
Pavitha, N., Ratnaparkhi, P., Uzair, A., More, A., Raj, S., Yadav, P., 2022b. Explainable AI for sentiment analysis. In: ICT with Intelligent Applications: Proceedings of ICTIS 2022, vol. 1, Springer, pp. 429–439.
Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., Jia, W., Yu, S., 2022. A survey on deep learning for textual emotion analysis in social networks. Digit. Commun. Netw. 8 (5), 745–762.
Pérez-Rosas, V., Mihalcea, R., Morency, L.-P., 2013. Utterance-level multimodal sentiment analysis. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 973–982.
Peterson, L.E., 2009. K-nearest neighbor. Scholarpedia 4 (2), 1883.
Poria, S., Cambria, E., Howard, N., Huang, G.-B., Hussain, A., 2016. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing 174, 50–59.
Qorib, M., Oladunni, T., Denis, M., Ososanya, E., Cotae, P., 2023. COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Syst. Appl. 212, 118715.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al., 2019. Language models are unsupervised multitask learners. OpenAI blog 1 (8), 9.
Rahm, E., Do, H.H., et al., 2000. Data cleaning: Problems and current approaches. IEEE Data Eng. Bull. 23 (4), 3–13.
Rane, A., Kumar, A., 2018. Sentiment classification system of twitter data for US airline service analysis. In: 2018 IEEE 42nd Annual Computer Software and Applications Conference. COMPSAC, vol. 1, IEEE, pp. 769–773.
Reitmaier, T., Wallington, E., Kalarikalayil Raju, D., Klejch, O., Pearson, J., Jones, M., Bell, P., Robinson, S., 2022. Opportunities and challenges of automatic speech recognition systems for low-resource language speakers. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. pp. 1–17.
Revathy, G., Alghamdi, S.A., Alahmari, S.M., Yonbawi, S.R., Kumar, A., Haq, M.A., 2022. Sentiment analysis using machine learning: Progress in the machine intelligence for data science. Sustain. Energy Technol. Assess. 53, 102557.
Rish, I., et al., 2001. An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, (no. 22), pp. 41–46.
Saad, E., Din, S., Jamil, R., Rustam, F., Mehmood, A., Ashraf, I., Choi, G.S., 2021. Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums. IEEE Access 9, 85721–85737.
Sabour, S., Frosst, N., Hinton, G.E., 2017. Dynamic routing between capsules. In: Advances in Neural Information Processing Systems, vol. 30.
Sakhare, N.N., Shaik, I.S., 2023. Spatial federated learning approach for the sentiment analysis of stock news stored on blockchain. Spatial Inf. Res. 1–15.
Saranya, G., Geetha, G., Meenakshi, K., Karpagaselvi, S., et al., 2020. Sentiment analysis of healthcare tweets using SVM classifier. In: 2020 International Conference on Power, Energy, Control and Transmission Systems. ICPECTS, IEEE, pp. 1–3.
Shah, P., Swaminarayan, P., Patel, M., 2022. Sentiment analysis on film review in Gujarati language using machine learning. Int. J. Electr. Comput. Eng. 12 (1), 1030.
Shaik, A., Devi, B.A., Baskaran, R., Bojjawar, S., Vidyullatha, P., Balaji, P., 2023. Recurrent neural network with emperor penguin-based Salp swarm (RNN-EPS2) algorithm for emoji based sentiment analysis. Multimedia Tools Appl. 1–20.
Shanmugavadivel, K., Sampath, S.H., Nandhakumar, P., Mahalingam, P., Subramanian, M., Kumaresan, P.K., Priyadharshini, R., 2022. An analysis of machine learning models for sentiment analysis of tamil code-mixed data. Comput. Speech Lang. 76, 101407.
Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., Catanzaro, B., 2019. Megatron-lm: Training multi-billion parameter language models using model parallelism. arXiv preprint arXiv:1909.08053.
Siddique, M.M., Kumar, S., 2023. Sentiment analysis on educational tweets: A case of national education policy 2020. InC4, In: 2023 IEEE International Conference on Contemporary Computing and Communications, vol. 1, IEEE, pp. 1–6.
Singh, R., Singh, R., 2023. Applications of sentiment analysis and machine learning techniques in disease outbreak prediction–A review. Mater. Today: Proc. 81, 1006–1011.
Singh, P., Srivastava, G., Singh, S., Kumar, S., 2023. Intelligent movie recommender framework based on content-based & collaborative filtering assisted with sentiment analysis. Int. J. Adv. Res. Comput. Sci. 14 (3).
Smith, S., Patwary, M., Norick, B., LeGresley, P., Rajbhandari, S., Casper, J., Liu, Z., Prabhumoye, S., Zerveas, G., Korthikanti, V., et al., 2022. Using deepspeed and megatron to train megatron-turing nlg 530b, a large-scale generative language model. arXiv preprint arXiv:2201.11990.
Sobieszek, A., Price, T., 2022. Playing games with AIs: The limits of GPT-3 and similar large language models. Minds Mach. 32 (2), 341–364.
Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 1631–1642.
Storey, V.C., Park, E.H., 2022. An ontology of emotion process to support sentiment analysis. J. Assoc. Inf. Syst. 23 (4), 999–1036.
Sultana, N., Sultana, R., Rasel, R.I., Hoque, M.M., 2022. Aspect-based sentiment analysis of bangla comments on entertainment domain. In: 2022 25th International Conference on Computer and Information Technology. ICCIT, IEEE, pp. 953–958.
Sun, T., Ni, J., Wang, W., Jing, L., Wei, Y., Nie, L., 2023. General debiasing for multimodal sentiment analysis. arXiv preprint arXiv:2307.10511.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.
Tang, D., Qin, B., Liu, T., Yang, Y., 2015. User modeling with neural network for review rating prediction. In: Twenty-Fourth International Joint Conference on Artificial Intelligence.
Tarcar, A.K., Tiwari, A., Dhaimodker, V.N., Rebelo, P., Desai, R., Rao, D., 2019. Healthcare NER models using language model pretraining. arXiv preprint arXiv:1910.11241.
Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K., Gutierrez, L., Tan, T.F., Ting, D.S.W., 2023. Large language models in medicine. Nat. Med. 29 (8), 1930–1940.
Tubishat, M., Al-Obeidat, F., Shuhaiber, A., 2023. Sentiment analysis of using ChatGPT in education. In: 2023 International Conference on Smart Applications, Communications and Networking. SmartNets, IEEE, pp. 1–7.
Tunca, S., Sezen, B., Wilk, V., 2023. An exploratory content and sentiment analysis of the guardian metaverse articles using leximancer and natural language processing. J. Big Data 10 (1), 82.
Usart, M., Grimalt-Álvaro, C., Iglesias-Estradé, A.M., 2023. Gender-sensitive sentiment analysis for estimating the emotional climate in online teacher education. Learn. Environ. Res. 26 (1), 77–96.
Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., Sandoval-Almazán, R., 2022. Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. Cogn. Comput. 1–16.
Vatambeti, R., Mantena, S.V., Kiran, K., Manohar, M., Manjunath, C., 2023. Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique. Cluster Comput. 1–17.

Venkataraman, J., Mohandoss, L., 2023. FBO-RNN: Fuzzy butterfly optimization-based RNN-LSTM for extracting sentiments from Twitter Emoji database. Concurr. Comput.: Pract. Exper. 35 (12), e7683.

Vohra, A., Garg, R., 2023. Deep learning based sentiment analysis of public perception of working from home through tweets. J. Intell. Inf. Syst. 60 (1), 255–274.

Wang, M., Cao, D., Li, L., Li, S., Ji, R., 2014. Microblog sentiment analysis based on cross-media bag-of-words model. In: Proceedings of International Conference on Internet Multimedia Computing and Service. pp. 76–80.

Wankhade, M., Rao, A.C.S., Kulkarni, C., 2022. A survey on sentiment analysis methods, applications, and challenges. Artif. Intell. Rev. 55 (7), 5731–5780.

Wen, J., Jiang, D., Tu, G., Liu, C., Cambria, E., 2023. Dynamic interactive multiview memory network for emotion recognition in conversation. Inf. Fusion 91, 123–133.

Wu, T., Peng, J., Zhang, W., Zhang, H., Tan, S., Yi, F., Ma, C., Huang, Y., 2022. Video sentiment analysis with bimodal information-augmented multi-head attention. Knowl.-Based Syst. 235, 107676.

Xu, Q.A., Chang, V., Jayne, C., 2022. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decis. Anal. J. 3, 100073.

Xu, G., Meng, Y., Qiu, X., Yu, Z., Wu, X., 2019. Sentiment analysis of comment texts based on BiLSTM. Ieee Access 7, 51522–51532.

Yadav, N.B., 2023. Harnessing customer feedback for product recommendations: An aspect-level sentiment analysis framework. Human-Centric Intell. Syst. 3 (2), 57–67.

Yadav, S.K., Bhushan, M., Gupta, S., 2015. Multimodal sentiment analysis: Sentiment analysis using audiovisual format. In: 2015 2nd International Conference on Computing for Sustainable Global Development. Indiacom, IEEE, pp. 1415–1419.

Yang, L., Li, Y., Wang, J., Sherratt, R.S., 2020. Sentiment analysis for e-commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE Access 8, 23522–23530.

Yekrangi, M., Abdolvand, N., 2021. Financial markets sentiment analysis: Developing a specialized lexicon. J. Intell. Inf. Syst. 57, 127–146.

Yekrangi, M., Nikolov, N.S., 2023. Domain-specific sentiment analysis: An optimized deep learning approach for the financial markets. IEEE Access.

Yi, C., Wang, J., Cheng, N., Zhou, S., Xu, B., 2020. Applying wav2vec2. 0 to speech recognition in various low-resource languages. arXiv preprint arXiv:2012.12121.

Yin, J.Y.B., Saad, N.H.M., Yaacob, Z., 2022. Exploring sentiment analysis on e-commerce business: Lazada and shopee. Tem J. 11 (4), 1508–1519.

You, Q., Luo, J., Jin, H., Yang, J., 2015. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, (no. 1).

You, Q., Luo, J., Jin, H., Yang, J., 2016. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. pp. 13–22.

Yu, J., Chen, K., Xia, R., 2022. Hierarchical interactive multimodal transformer for aspect-based multimodal sentiment analysis. IEEE Trans. Affect. Comput..

Zadeh, A., Zellers, R., Pincus, E., Morency, L.-P., 2016. Mosi: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. arXiv preprint arXiv:1606.06259.

Zhai, G., Yang, Y., Wang, H., Du, S., 2020. Multi-attention fusion modeling for sentiment analysis of educational big data. Big Data Min. Anal. 3 (4), 311–319.

Zhang, L., Wang, S., Liu, B., 2018. Deep learning for sentiment analysis: A survey. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 8 (4), e1253.

Zhang, Y., Wang, M., Tiwari, P., Li, Q., Wang, B., Qin, J., 2023a. DialogueLLM: Context and emotion knowledge-tuned LLaMA models for emotion recognition in conversations. arXiv preprint arXiv:2310.11374.

Zhang, Y., Yu, Y., Guo, Q., Wang, B., Zhao, D., Uprety, S., Song, D., Li, Q., Qin, J., 2023b. CMMA: Benchmarking multi-affection detection in Chinese multi-modal conversations. In: Thirty-Seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track.

Zhang, S., Zhang, D., Zhong, H., Wang, G., 2020a. A multiclassification model of sentiment for e-commerce reviews. IEEE Access 8, 189513–189526.

Zhang, Y., Zhao, Z., Wang, P., Li, X., Rong, L., Song, D., 2020b. ScenarioSA: A dyadic conversational database for interactive sentiment analysis. IEEE Access 8, 90652–90664.

Zhou, J., Tian, J., Wang, R., Wu, Y., Xiao, W., He, L., 2020. Sentix: A sentiment-aware pre-trained model for cross-domain sentiment analysis. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 568–579.

Zhou, J., Ye, J.-m., 2023. Sentiment analysis in education research: A review of journal publications. Interact. Learn. Environ. 31 (3), 1252–1264.

Zuo, E., Zhao, H., Chen, B., Chen, Q., 2020. Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. IEEE Access 8, 37967–37975. http://dx.doi.org/10.1109/ACCESS.2020.2975244.