International Journal of Basic and Applied Sciences, 14 (2) (2025) 34-48

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/mhv83077
Research paper

Review of sentiment analysis in social media using big data: techniques, tools, and frameworks

Srishti Sudhir Patil [1] *, Vaishali Pankaj Suryawanshi [2], Sudhir Madhav Patil [3], Sheetal Pravin Girase [2], Dhananjay Abhilash Bhagat [2]

[1] Student, Third Year B.Tech.: Computer Science and Engineering, Department of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University (MIT-WPU), Kothrud, Pune: 411038, Maharashtra State, India
[2] Assistant Professor, Department of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University (MIT-WPU), Kothrud, Pune: 411038, Maharashtra State, India
[3] Associate Professor, Department of Manufacturing Engineering and Industrial Management, COEP Technological University (COEP Tech), Chhatrapati Shivajinagar, Pune: 411005, Maharashtra State, India
**Corresponding author E-mail: 1032220293@mitwpu.edu.in*

Received: April 14, 2025, Accepted: June 3, 2025, Published: June 9, 2025

Abstract

Sentiment analysis on social media has emerged as a vital research area due to the growing volume of user-generated content and the increasing reliance on data-driven decision-making. The adoption of big data technologies has greatly improved sentiment analysis by enabling the rapid processing of unstructured big data. This review presents an in-depth analysis of sentiment analysis methodologies, covering both conventional machine learning (ML) techniques-such as Naïve Bayes, Support Vector Machines, Decision Trees, and Random Forest and advanced deep learning (DL) models, including Recurrent Neural Networks, Long Short-Term Memory Networks, Convolutional Neu-ral Networks, and Transformer-based architectures like Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformers (GPT). Furthermore, it examines big data frameworks like Hadoop, Apache Spark, and Apache Flink, along with Natural Language Processing (NLP) tools such as the Natural Language Toolkit (NLTK), spaCy, TextBlob, and Stanford NLP. The paper also discusses ML/DL frameworks like Scikit-learn, TensorFlow, PyTorch, and Keras, along with cloud and edge computing solutions like Google Cloud Artificial Intelligence (AI), Amazon Web Services (AWS) Comprehend, and Edge AI (NVIDIA Jetson). Despite technological advancements, several challenges persist, including issues related to data quality, real-time processing limitations, multilingual analysis complexities, and ethical concerns regarding bias and privacy. The field is also witnessing promising developments, such as Explainable Artificial Intelligence (XAI), federated learning, edge computing, and quantum computing, which offer new directions for future research and practical implementations. This review provides researchers and professionals with valuable insights, outlining potential improvements in sentiment analysis techniques to enhance accuracy, scalability, and ethical considerations across various sectors, including business, healthcare, and smart manufacturing.

Keywords: *Big Data; Deep Learning; Machine Learning; Natural Language Processing; Sentiment Analysis; Social Media.*

1. Introduction

The growing volume of user-generated content on social media has made sentiment analysis a critical area of research [1], [2]. By utilizing big data, sentiment analysis enables real-time assessment of public opinion [1], influencing business strategies [3], policy decisions [4], and social trends [1], [3]. This review examines the key methodologies, tools, and frameworks essential for effective sentiment analysis.

1.1. Background and importance of sentiment analysis in social media

Social media platforms generate an immense amount of text-based data daily, capturing user sentiments on diverse topics. Analyzing these sentiments allows organizations to gain insights into consumer behavior, anticipate market trends, and monitor social movements [1], [5]. Businesses employ sentiment analysis for brand reputation management [6], while policymakers leverage it to gauge public sentiment and inform decision-making [7].

1.2. Role of big data in analyzing sentiment trends

Big data plays a crucial role in processing the vast and unstructured information generated on social media [8]. Advanced computational techniques, such as distributed computing and AI-driven models, enable efficient extraction of meaningful insights. Effective data management facilitates timely decision-making across various industries [9].

1.3. Objectives and scope of this review

This review provides a comprehensive examination of sentiment analysis techniques, with a particular focus on big data methodologies and tools. It explores machine learning and deep learning approaches, along with their real-world applications. Additionally, the review discusses existing challenges, ethical considerations, and future directions for enhancing sentiment analysis capabilities.

1.4. Paper organization

The paper is structured as follows: Section 2 discusses fundamental concepts and sentiment classification levels. Section 3 delves into big data sources and associated challenges. Section 4 presents sentiment analysis workflow in social media. Section 5 discusses the evolution of sentiment analysis techniques. Section 6 examines machine learning and deep learning techniques. Section 7 presents various big data sentiment analysis tools and frameworks. Section 8 outlines existing challenges, while Section 9 explores future research opportunities. Finally, Section 10 provides concluding remarks.

2. Fundamentals of sentiment analysis

Sentiment analysis, also referred to as opinion mining, involves extracting and interpreting subjective information from textual data. It plays a vital role in assessing user opinions, emotions, and attitudes across different fields [10]. This section provides an overview of key concepts, sentiment classification levels, and various methodological approaches used in sentiment analysis.

2.1. Definition and core concepts

Sentiment analysis is the process of evaluating text to determine its underlying emotional tone, typically classified as positive, negative, or neutral [10]. It integrates techniques from natural language processing (NLP) [1], computational linguistics [11], and machine learning [12] to derive meaningful insights. Its applications span across multiple domains, including social media analysis [13], market research [14], and customer feedback assessment [15].

2.2. Levels of sentiment classification

Sentiment analysis can be conducted at different levels of granularity to capture varying sentiment expressions within a text.

2.2.1. Document-level sentiment analysis

This approach evaluates the overall sentiment of an entire document or post, if the text conveys a single dominant sentiment. It is particularly useful for analyzing reviews, opinion articles, and other long-form content [16].

2.2.2. Sentence-level sentiment analysis

At this level, sentiment is determined on a per-sentence basis, identifying whether a given sentence expresses a positive, negative, or neutral sentiment. This method is beneficial for texts containing multiple viewpoints, allowing for a more detailed sentiment breakdown [16].

2.2.3. Phrase-level sentiment analysis

Phrase-level sentiment analysis examines sentiment within individual phrases rather than entire documents or sentences, providing a more nuanced understanding. It captures context-sensitive expressions, enabling multi-aspect analysis in reviews and social media. Compared to document- and sentence-level analysis, it offers finer sentiment granularity, making it valuable for product reviews, market research, and social media monitoring [17], [18].

2.2.4. Aspect-based sentiment analysis

Aspect-based sentiment analysis (ABSA) focuses on identifying sentiments associated with specific aspects or features within a text. For instance, in a product review, a user might express satisfaction with battery life but dissatisfaction with camera quality. ABSA helps in gaining a more nuanced understanding of opinions related to different attributes [19].

2.2.5. Entity-level sentiment analysis

Entity-level sentiment analysis (ELSA) focuses on identifying sentiment toward specific entities within social media data. Using big data techniques, it extracts opinions about brands, products, or individuals. Machine learning and NLP tools analyze sentiment at a granular level, enabling precise insights for businesses, policymakers, and researchers in decision-making [20].

2.2.6. Comparative sentiment analysis

Comparative sentiment analysis (CSA) evaluates sentiment differences between multiple entities, topics, or timeframes in social media data. It requires comparative phrase detection and may struggle with context-dependent comparisons. Leveraging big data tools, it identifies trends, preferences, and competitive insights. Machine learning and NLP frameworks enhance accuracy, enabling businesses and researchers to assess public opinion shifts and make data-driven decisions effectively [21].

2.2.7. Topic-level sentiment analysis

Topic-level sentiment analysis focuses on identifying sentiment associated with specific topics within a dataset. Using NLP and machine learning, it categorizes opinions expressed in social media, reviews, or discussions, helping businesses and researchers understand public perception, emerging trends, and sentiment variations across different subject areas [22].

2.3. Types/approaches of sentiment analysis

Various techniques are employed in sentiment analysis, each with its strengths and limitations.

2.3.1. Lexicon-based methods

Lexicon-based approaches utilize predefined sentiment dictionaries that contain words associated with positive, negative, or neutral emotions. Sentiment scores are assigned based on word occurrences within a given text [23]. While these methods are effective for basic sentiment identification, they often struggle with understanding contextual meaning and detecting sarcasm [24].

2.3.2. Machine learning-based methods

Machine learning techniques apply supervised or unsupervised algorithms to classify sentiment. Common models include Naïve Bayes, Support Vector Machines (SVM), and deep learning architectures such as Long Short-Term Memory (LSTM) networks and transformer-based models [25]. These approaches generally provide higher accuracy but require extensive labeled datasets for effective training [26].

2.3.3. Hybrid approaches

Hybrid models integrate lexicon-based techniques with machine learning algorithms to improve sentiment classification. By combining linguistic rules with the adaptability of machine learning models, these approaches offer a more comprehensive and reliable analysis, particularly for handling complex sentiment expressions [27].

3. Big data in sentiment analysis

Social media produces a massive amount of user-generated content, making big data an essential element in sentiment analysis. Effectively processing, analyzing, and deriving valuable insights from this vast data pool is crucial for businesses, policymakers, and researchers [1]. This section explores the key characteristics of big data, its sources, challenges associated with handling big sentiment data, and preprocessing techniques used to improve the accuracy of sentiment analysis.

3.1. Characteristics of big data in social media sentiment analysis

Big data in sentiment analysis is defined by several key attributes as presented further. Volume - social media platforms generate an extensive amount of text, images, and videos daily. Velocity - data is produced in real-time, necessitating fast processing for timely insights. Variety - content appears in multiple formats, including text, emojis, audio, and video. Veracity - user-generated content can be misleading or inaccurate, posing challenges for data reliability. Value - Extracting meaningful sentiment patterns from big data is essential for informed decision-making [3], [28]. Given the dynamic nature of social media, advanced computational techniques are required to process and analyze sentiment trends effectively. While big data provides valuable insights, it also presents challenges in terms of scale, complexity, and accuracy [1], [7], [9].

3.2. Sources of big data in sentiment analysis

Social media platforms are a primary source of big data for sentiment analysis, offering diverse user-generated content that reflects opinions, trends, and public sentiment [1], [7]. Data is collected through APIs [29], web scraping [30], [31], and user interactions [7], forming the foundation for sentiment classification.

X (formerly Twitter): A key platform for sentiment analysis, offering real-time, concise textual data on trending topics. Reddit: Provides structured discussion threads with detailed user opinions. YouTube: Video metadata and comment sections reveal sentiment trends. Facebook: User posts, comments, and reactions provide valuable sentiment insights. Instagram: Image captions, hashtags, and comments reflect public sentiment. LinkedIn: Professional discussions and job trends contribute to sentiment analysis. WhatsApp: Group chats and shared media influence sentiment patterns. TikTok: Short-form videos and user comments reveal engagement trends. WeChat: Texts, media sharing, and public accounts provide sentiment data. Telegram: Public channels and group interactions offer sentiment insights. Snapchat: Ephemeral content and reactions help gauge user sentiment. Pinterest: Pin descriptions and comments indicate user interests. Discord: Server chats and community discussions reflect opinions. Twitch: Live chat interactions capture real-time audience sentiment. Tencent (QQ): Social interactions and multimedia sharing influence sentiment analysis. Messenger: Private and group conversations contribute to sentiment trends. Tumblr: User blogs, multimedia content, and hashtags reflect niche community sentiment trends. Vimeo: Video comments and engagement metrics provide sentiment insights for creative industries [1], [3], [7], [8].

Each platform offers unique sentiment patterns, and combining multiple sources enhances the accuracy of sentiment analysis.

3.3. Challenges in managing big sentiment data

Analyzing sentiment data from social media involves several challenges, as outlined further. Data Overload: The vast volume of social media data demands scalable storage solutions and high-performance computational resources [32]. Noise and Irrelevant Content: Spam, bot-generated posts, and informal language can distort sentiment analysis results [33]. Real-Time Processing: Extracting insights from continuously evolving data streams requires efficient and high-speed analytical tools [34], [35]. Multilingual Complexity: Sentiment analysis must account for multiple languages, dialects, and cultural nuances [36]. Privacy and Ethical Issues: Handling personal data raises concerns about compliance with regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) [35], [37].

To address these challenges, advanced data management strategies, AI-driven filtering techniques, and scalable computing infrastructures are essential [35], [38].

3.4. Preprocessing techniques for big data sentiment analysis

Effective preprocessing is crucial for refining big sentiment data, ensuring higher accuracy and efficiency in analysis [8], [39], [40]. The key techniques are discussed further. Data Cleaning: Removes duplicate entries, irrelevant content, and special characters [41], [42]. Tokenization: Breaks text into smaller units, such as words or phrases, for analysis [41]. Stopword Removal: Filters out frequently used words (e.g., "the," "is") that do not contribute to sentiment detection [42]. Stemming and Lemmatization: Converts words into their base forms to maintain consistency in analysis [41], [42]. Emoji and Slang Processing: Transforms emojis and informal expressions into standardized sentiment-bearing terms [43]. Feature Extraction: Uses methods like fastText, Term Frequency-Inverse Document Frequency (TF-IDF), Word2vec, word embeddings, global vectors (GloVe), and sentiment lexicons to enhance text representation [43]. Normalization: Scales data to a specific range, improving model performance and training speed [44]. Data Balancing: Adjusts class distribution to prevent biased model learning in imbalanced datasets [44]. Standardization: Transforms features to have zero mean and unit variance for consistent data scaling [45].

These preprocessing techniques help eliminate noise, improve sentiment classification, and optimize the efficiency of big data sentiment analysis.

4. Sentiment analysis workflow in social media

This section presents the standard workflow followed in sentiment analysis of social media data using big data approaches. It highlights the key stages, including data acquisition, preprocessing, sentiment detection, and result interpretation. Each step plays a vital role in managing large-scale, unstructured data from social platforms and transforming it into actionable insights. A clear understanding of this workflow supports the development of robust and scalable sentiment analysis systems. Figure 1 depicts a typical workflow commonly employed in analyzing sentiments expressed through social media. Table 1 presents the detailed workflow of sentiment analysis in social media using big data along with the objective, methods, tools, techniques, and frameworks for each step.

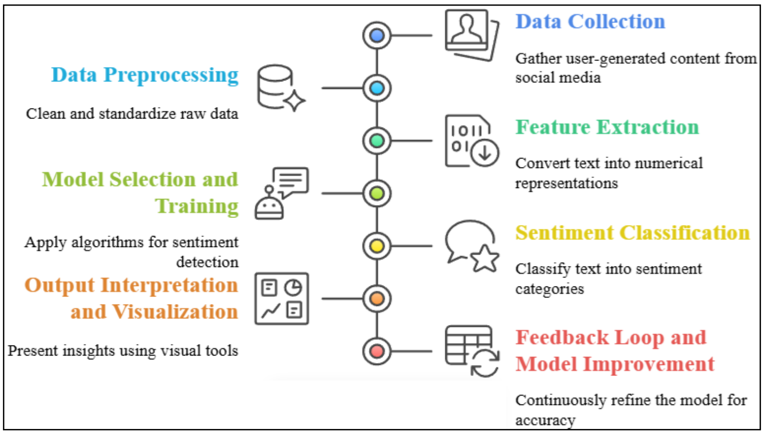

Fig. 1: Sentiment Analysis Workflow in Social Media.

Table 1: Detailed Workflow with Stepwise Objective, Tools, Techniques and Frameworks for Sentiment Analysis in Social Media Using Big Data

Step	Objective	Description	Tools/Techniques/Frameworks	
Data Collection	Gather large volumes of user-generated content from various social media platforms.	Sources: • Twitter (tweets, hashtags), • Facebook (comments, posts), • Reddit, • Instagram, • YouTube, • blogs, forum	Tools/Technologies: ~ APIs: Twitter API, Facebook Graph API, Reddit API ~ Web Scraping: BeautifulSoup, Scrapy, Selenium ~ Streaming Platforms: Apache Kafka, Apache Flume for real-time data ingestion	
Data Preprocessing	Clean and standardize raw data to prepare it for analysis.	Key Steps: • Text cleaning (removing URLs, mentions, hashtags, emojis, special characters), • Tokenization, • Stopword removal, • Lemmatization/Stemming, • Language detection and filtering, • Handling misspellings, slang, and abbreviations, • Dealing with code-mixed or multilingual data	Tools: ~ NLTK, ~ SpaCy, ~ Gensim, ~ LangDetect, ~ Polyglot (for language detection) ~ Regex, ~ custom rule-based scripts	Big Data Frameworks: ~ Apache Spark (using MLlib and Spark NLP), ~ Hadoop (for distributed data storage and cleaning), ~ HDFS
Feature Extraction	Convert textual data into numer-		Tools & Libraries: ~ Scikit-learn (TF-IDF)	Techniques: ~ Bag of Words (BoW) ~ TF-IDF

	ical representations suitable for model input.		~ Gensim (Word2Vec, FastText) ~ Hugging Face Transformers (BERT-based models) ~ Spark MLlib, ~ TensorFlow on Spark ~ TensorFlow Hub / PyTorch with pretrained embeddings	~ Word Embeddings (Word2Vec, GloVe, FastText) ~ Contextual Embeddings (BERT, RoBERTa, mBERT for multilingual support)	
Model Selection and Training	Apply suitable algorithms to detect sentiment polarity or emotion categories.	Approaches: • Traditional ML: ~ Naïve Bayes, ~ SVM, ~ Logistic Regression, ~ Random Forest • Deep Learning: ~ CNNs, ~ LSTMs, ~ GRUs, ~ BiLSTMs • Transformers: ~ BERT, ~ RoBERTa, ~ XLM-RoBERTa • Ensemble Methods: ~ Voting, ~ Bagging, ~ Stacking classifiers	Tools/Frameworks: ~ Scikit-learn (ML models), ~ Keras, ~ TensorFlow, ~ PyTorch (DL models), ~ Hugging Face Transformers ~ AutoML tools: Google AutoML, H2O.ai	Techniques: ~ Docker, ~ Kubernetes, ~ REST APIs	Big Data Integration: ~ Apache Spark MLlib ~ TensorFlowOnSpark, Horovod for distributed deep learning
Sentiment Classification	Classify text into sentiment categories.	Typical Output Classes: • Polarity: ~ Positive, ~ Negative, ~ Neutral • Emotions: ~ Joy, ~ Sadness, ~ Anger, ~ Fear, etc.	Tools: ~ VADER (Valence Aware Dictionary and sEntiment Reasoner), ~ Lexalytics ~ Brandwatch ~ Talkwalker ~ Sentiment.ai ~ WEKA ~ Python ~ R	Methods: • Lexicon-based approaches • Rule-based approaches • Machine learning models: ~ Naive Bayes ~ Support Vector Machines (SVM) ~ Logistic Regression ~ Random Forest ~ Gradient Boosting Machines (GBM) • Deep learning models: ~ Recurrent Neural Networks (RNN) ~ Long Short-Term Memory (LSTM) ~ Gated Recurrent Unit (GRU)	Considerations: ~ Multilingual sentiment handling ~ Cultural sentiment nuances ~ Handling sarcasm and irony
Output Interpretation and Visualization	Present actionable insights to stakeholders using visual tools and dashboards.	Methods: • Sentiment distribution charts • Time-series sentiment trends • Geo-location based sentiment mapping • Word clouds or topic clustering	Tools: ~ Tableau, ~ Power BI ~ Kibana (for Elastic Stack users) ~ D3.js, ~ Plotly, ~ Matplotlib, ~ Seaborn ~ Dash (for interactive Python dashboards)		Techniques: ~ Docker, ~ Kubernetes, ~ REST APIs
Feedback Loop and Model Improvement	Continuously refine the model for improved accuracy.	Processes: • Error analysis • User feedback incorporation • Active learning • Retraining with updated datasets	Tools: ~ MLflow, Weights & Biases (for model tracking) ~ A/B testing frameworks		Techniques: ~ Docker, ~ Kubernetes, ~ REST APIs

5. Evolution of sentiment analysis techniques

This section outlines the evolution of sentiment analysis techniques, tracing their progression from traditional methods to modern big data-driven approaches. It highlights how advancements in machine learning, natural language processing, and deep learning have transformed sentiment analysis, enabling more accurate and context-aware insights from social media data. Understanding this evolution provides valuable context for selecting appropriate techniques based on data scale, complexity, and application requirements. Figure 2 illustrates the key stages in the development of sentiment analysis methods over time. Table 2 presents the detailed timeline in sentiment analysis along with the key focus and their respective tools, techniques, and frameworks.

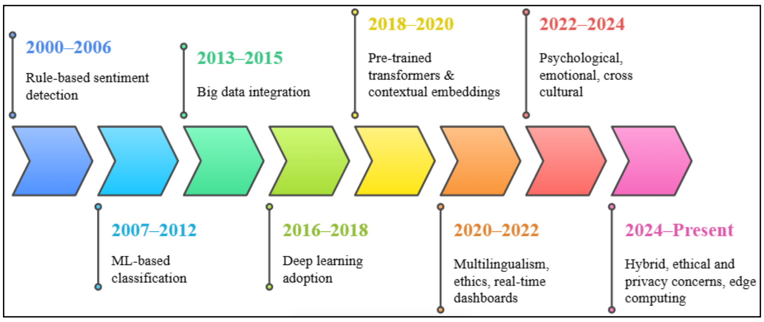

Fig. 2: Timeline Showing Evolution of Sentiment Analysis.

Table 2: Detailed Timeline in Sentiment Analysis Along with the Key Focus and Their Respective Tools, Techniques and Frameworks

Period	Key Focus	Techniques	Tools/Frameworks	Big Data Relevance
2000–2006	Rule-based sentiment detection	Rule-based, Lexicons	SentiWordNet, Stanford NLP, LIWC	Minimal
2007–2012	ML-based classification	ML Classifiers	Scikit-learn, Weka, NLTK, Naïve Bayes, SVM	Early API usage
2013–2015	Big data integration	Word Embeddings, LSTMs	Gensim, TensorFlow, Keras	Hadoop, Spark for batch jobs, Kafka, Flume
2016–2018	Deep learning adoption	Attention, BiLSTM, Real-time	PyTorch, Kafka, Spark MLlib	Streaming integration (Kafka, Flink)
2018–2020	Pre-trained transformers & contextual embeddings	Transformers (BERT)	Hugging Face, TensorFlow Hub	GPU/TPU-based distributed training
2020–2022	Multilingualism, ethics, and real-time dashboards	Multilingual Models	XLM-R, mBERT, IndicNLP, AutoML, real-time visualization tools	Cross-lingual datasets, federated learning
2022–2024	Psychological, emotional, cross-cultural	XAI, Emotion-aware DL	LIME, SHAP, GPT APIs	MLOps, Cloud-native platforms
2024–Present	Hybrid, ethical, and privacy concerns, edge computing	Foundation Models, Multimodal	LangChain, Haystack, LLMOps	Real-time cross-modal sentiment at scale

6. Machine learning and deep learning techniques for sentiment analysis

Sentiment analysis utilizes both traditional machine learning (ML) and advanced deep learning (DL) models to interpret emotions in textual data. While ML approaches depend on statistical techniques and predefined features, DL methods employ hierarchical learning for improved accuracy [3], [9]. This section explores key ML and DL techniques and evaluates their effectiveness in sentiment classification.

6.1. Traditional machine learning methods

ML-based sentiment analysis relies on feature engineering and statistical classification algorithms [8], [26]. Some commonly used techniques are discussed further. Naïve Bayes (NB): A probabilistic model based on Bayes' theorem that assumes feature independence. It is efficient for text classification but struggles with complex sentence structures [25]. Support Vector Machines (SVM): A supervised learning technique that identifies an optimal hyperplane for classification. It performs well on high-dimensional data but requires careful feature selection [25], [46]. Decision Trees (DT): A rule-based classifier that segments data based on feature values. While interpretable, it is prone to overfitting, especially in large datasets [25]. Random Forest (RF): An ensemble of decision trees that improves classification performance by reducing overfitting, making it robust for text-based sentiment analysis [46].

These ML approaches require manual feature extraction techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) or word embeddings to convert text into structured data for analysis [47].

6.2. Deep learning approaches

Deep learning models enhance sentiment analysis by automatically extracting complex features from text data [9], [22], [25], [26]. Key approaches are discussed further. Recurrent Neural Networks (RNNs): Designed for sequential data, RNNs capture contextual dependencies but suffer from vanishing gradient issues in long texts [9], [25], [48]. Long Short-Term Memory (LSTM): A type of RNN that incorporates memory cells to preserve long-term dependencies, improving the analysis of complex sentence structures [48]. Convolutional Neural Networks (CNNs): Originally used for image processing, CNNs can capture local text patterns and are effective for short text sentiment analysis [25]. Transformer Models include BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). BERT excels in understanding context by processing text bidirectionally. GPT uses an autoregressive approach for text generation and sentiment classification [44], [45].

Deep learning eliminates the need for manual feature extraction and significantly enhances sentiment classification accuracy [8], [9], [26].

6.3. Ensemble learning approaches

Ensemble learning techniques play a crucial role in sentiment analysis by combining multiple models to enhance accuracy, robustness, and generalization [9], [49]. One common approach is bagging, where multiple classifiers are trained on different subsets of the data, and their predictions are aggregated to reduce overfitting and variance. Random Forest, a popular bagging-based method, utilizes multiple decision trees to achieve stable and accurate sentiment classification. Boosting methods, such as AdaBoost, Gradient Boosting, and XGBoost, sequentially train weak models, with each iteration focusing on errors from the previous one, thereby improving prediction performance by reducing bias and variance [50-53].

Stacking is another ensemble technique where diverse models, including machine learning and deep learning classifiers, are combined using a meta-learner, which further refines the final prediction. Voting ensembles aggregate predictions from multiple models through majority voting or weighted averaging, leveraging the strengths of different algorithms. Blending, a variation of stacking, uses a separate validation set to train the meta-learner, mitigating overfitting risks [52], [53]. These ensemble methods are widely used in sentiment analysis applications, including social media monitoring, product reviews, and customer feedback assessment, ensuring more reliable sentiment classification [9], [49-53].

6.4. Comparative analysis of ML, DL, and ensemble learning for sentiment analysis

Among the various computational techniques employed for sentiment analysis, ML, DL, and Ensemble Learning stand out due to their efficiency and scalability. Each of these approaches has its advantages and limitations when applied to sentiment analysis in social media big datasets [45], [54], [55].

ML techniques, such as SVM, Naïve Bayes, and Random Forest, have been widely used for sentiment classification. These models rely on handcrafted features such as TF-IDF, n-grams, and lexicon-based features [25], [56-58]. ML algorithms require extensive feature engineering and domain-specific tuning to achieve optimal performance. However, their efficiency in handling moderate-sized datasets makes them suitable for applications where computational resources are limited [26], [59], [60]. Traditional ML models are interpretable and computationally efficient but struggle with context-dependent sentiment expressions and the complex linguistic structures found in social media posts [44], [45], [61].

DL, on the other hand, has revolutionized sentiment analysis by eliminating the need for manual feature engineering. DL models such as CNNs, RNNs, LSTMs, and Transformer-based architectures like BERT and GPT excel in capturing contextual information from text [9], [62]. These models can process massive social media datasets and extract intricate patterns in sentiment expressions. DL-based sentiment analysis benefits from word embeddings like Word2Vec, GloVe, and contextual embeddings, which significantly improve accuracy by preserving semantic meaning [8], [9], [63]. However, DL models require large amounts of labeled training data and high computational resources for training, making them less feasible for real-time applications on resource-constrained systems. Additionally, the black-box nature of DL models poses challenges in interpretability, limiting their adoption in applications requiring explainability [45], [64].

Ensemble Learning combines multiple base models to enhance prediction accuracy and robustness. Techniques such as Bagging, Boosting (e.g., AdaBoost, XGBoost), and Stacking integrate the strengths of different classifiers to achieve superior sentiment classification performance [53], [65], [66]. In the context of social media sentiment analysis, ensemble models can combine ML and DL approaches to balance interpretability and accuracy. For example, hybrid models using SVM with LSTM or Random Forest with BERT embeddings have demonstrated improved performance in handling noisy and imbalanced datasets [27], [61], [67]. Ensemble learning reduces overfitting and enhances generalization, making it suitable for large-scale social media sentiment analysis. However, the increased computational complexity and the need for hyperparameter tuning can make ensemble approaches challenging to deploy in real-time applications [52], [68], [69].

ML techniques provide computational efficiency and interpretability but require extensive feature engineering. DL models excel in contextual understanding but demand high computational resources and large datasets. Ensemble learning offers a balanced approach by leveraging multiple models to improve accuracy and robustness, albeit at the cost of increased complexity. The choice among ML, DL, and ensemble learning for sentiment analysis in social media using big data depends on factors such as dataset size, computational constraints, and the need for interpretability. Future advancements in hybrid models and transfer learning are likely to further enhance the effectiveness of these techniques for sentiment analysis in dynamic and large-scale social media environments. Table 3 presents the comparative analysis of ML, DL, and ensemble learning for sentiment analysis in social media using big data.

Table 3: Comparison of ML, DL and Ensemble Learning for Sentiment Analysis

Aspect	Machine Learning (ML)	Deep Learning (DL)	Ensemble Learning
Definition	Uses statistical and rule-based models to classify sentiment based on extracted features.	Uses artificial neural networks (ANNs) to automatically learn patterns and context in sentiment-related data.	Combines multiple models (ML, DL, or hybrid) to improve accuracy and robustness in sentiment classification.
Examples of Models	Naïve Bayes, Support Vector Machines (SVM), Random Forest, Logistic Regression.	Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Transformers (BERT, GPT).	Bagging (Random Forest), Boosting (AdaBoost, XGBoost), Stacking (combining ML and DL models).
Feature Engineering	Requires manual feature extraction (TF-IDF, n-grams, lexicon-based methods).	No manual feature extraction required; learns representations using embeddings (Word2Vec, GloVe, contextual embeddings).	May combine handcrafted and learned features depending on the base models used.
Performance on Large Datasets	Struggles with very large datasets; scalability is limited by feature engineering.	Performs well on large datasets, especially with GPU acceleration.	Improves performance on large datasets by combining strengths of multiple models.
Contextual Understanding	Limited ability to capture complex language nuances, sarcasm, and slang.	Excellent at understanding context, syntax, and semantics in sentiment analysis.	Can improve contextual understanding by integrating multiple models (e.g., combining SVM with BERT).
Computational Requirements	Low to moderate; can run efficiently on CPUs.	High; requires GPUs or TPUs for large-scale training and inference.	High; ensemble methods increase computational complexity due to multiple model training.
Training Data Requirements	Works well with small to moderately sized datasets.	Requires large labeled datasets for optimal performance.	May require a diverse dataset for training multiple models effectively.
Accuracy	Moderate; depends on the quality of feature engineering and model tuning.	High can outperform ML models by capturing deeper patterns in sentiment expressions.	Very high; leveraging multiple models improves generalization and accuracy.
Interpretability	High models like decision trees and SVMs offer transparency in classification.	Low-depth networks function as black-box models, making explainability difficult.	Moderate; some ensemble methods improve interpretability (e.g., Random Forest in bagging).
Handling of Noisy Data	Prone to misclassification due to reliance on handcrafted features.	Can learn from noisy data but may overfit if not properly regularized.	More robust to noise as different models compensate for each other's weaknesses.
Handling of Imbalanced Data	May require oversampling, under sampling, or cost-sensitive learning techniques.	Struggles with highly imbalanced data; class weights or synthetic data generation may be needed.	Performs better on imbalanced datasets by leveraging multiple models and balancing mechanisms.
Suitability for Real-time Applications	Suitable for real-time analysis due to lower computational overhead.	Challenging for real-time applications due to complex architectures and inference time.	Less suitable for real-time use due to increased computational demands.

Deployment Complexity	Easy to deploy; requires minimal infrastructure.	Complex to deploy; requires model optimization, GPUs, and cloud-based solutions.	More complex due to multiple models running in parallel or sequence.
Best Use Cases in Sentiment Analysis	Suitable for small-scale sentiment analysis, where interpretability and lower computational cost are preferred.	Ideal for large-scale social media sentiment analysis, especially for tasks requiring deep contextual understanding.	Best for improving accuracy in challenging sentiment analysis tasks, such as handling sarcasm, domain-specific language, and large imbalanced datasets.

7. Tools and frameworks for big data sentiment analysis

Effective sentiment analysis in social media depends on robust tools and frameworks for processing large datasets, natural language processing, and model training [47], [70], [71]. This section outlines key technologies used across different categories.

7.1. Big data processing tools

Handling big sentiment data requires scalable big data frameworks as discussed further [3], [8]. Hadoop: A distributed computing framework that processes vast datasets using the MapReduce model. It is effective for batch processing but lacks real-time capabilities [72], [73]. Apache Spark: An in-memory data processing engine supporting large-scale analytics, with its MLlib library improving sentiment classification performance [74]. Apache Flink: Designed for real-time data streaming, Flink enables dynamic tracking of sentiment trends on social media [75], [76]. These tools ensure efficient processing of large volumes of social media sentiment data [75-77].

7.2. Natural language processing libraries

NLP libraries support text preprocessing, feature extraction, and sentiment classification [78], [79]. Common libraries included in this are discussed further. NLTK (Natural Language Toolkit): A Python-based library offering techniques like tokenization, stemming, and sentiment analysis [80]. spaCy: A high-performance NLP library optimized for deep learning applications, offering fast tokenization and named entity recognition (NER) [81]. TextBlob: A user-friendly NLP library built on NLTK that simplifies sentiment classification and text analysis [82]. Stanford NLP: A toolkit developed by Stanford University, featuring advanced techniques such as dependency parsing and sentiment analysis [83]. These libraries enhance text processing capabilities, improving sentiment analysis accuracy.

7.3. Machine learning and deep learning frameworks

ML and DL frameworks provide essential tools for training and deploying sentiment analysis models. Key frameworks are discussed further. Scikit-learn: A versatile ML library offering Naïve Bayes, SVM, and Random Forest algorithms for sentiment classification [84]. TensorFlow: A powerful DL framework supporting neural network-based sentiment analysis, including transformer models like BERT [85]. PyTorch: A widely used deep learning framework that enables flexible model implementation for sentiment analysis [27], [86]. Keras: A high-level API built on TensorFlow, simplifying deep learning model development for text classification tasks [87].
These frameworks offer flexibility and scalability in implementing sentiment analysis models.

7.4. Cloud and edge computing solutions for sentiment analysis

Cloud and edge computing platforms enhance sentiment analysis by providing scalable infrastructure and real-time processing capabilities [3], [28]. Notable solutions are discussed further. Google Cloud AI: Offers pre-trained NLP models and scalable infrastructure for sentiment classification [88]. Amazon Web Services (AWS) Comprehend: A cloud-based NLP service supporting sentiment detection and text analysis [89]. Microsoft Azure Text Analytics: Provides AI-powered sentiment classification with robust language processing capabilities [89], [90]. Edge AI (NVIDIA Jetson, Intel Movidius): Enables real-time sentiment analysis at the edge, improving response time while enhancing privacy [91]. By leveraging these frameworks, organizations can efficiently analyze big sentiment data, gain valuable insights, and make informed decisions. Table 4 provides a comparison of different tools based on efficiency, scalability, and usability. Each tool offers unique advantages based on project requirements, with big data frameworks ensuring scalability, NLP libraries enhancing text processing, and ML/DL frameworks optimizing model accuracy.

Table 4: Comparison of Different Tools and Frameworks for Big Data Sentiment Analysis Based on Efficiency, Scalability, and Usability

Category	Tool/Framework	Efficiency	Scalability	Usability
Big Data Processing	Hadoop	High for batch processing	Scalable but slower than Spark	Moderate
	Apache Spark	High	Highly scalable	Moderate
	Apache Flink	Very high for real-time data	Scalable	Moderate
NLP Libraries	NLTK	Moderate	Limited scalability	Easy to use
	spaCy	High	Scalable	Easy to use
	TextBlob	Moderate	Limited scalability	Very easy to use
ML/DL Frameworks	Scikit-learn	High for ML models	Scalable for structured data	Easy to use
	TensorFlow	Very high	Highly scalable	Requires expertise
	PyTorch	Very high	Highly scalable	Flexible but complex
Cloud & Edge Computing	Google Cloud AI	High	Highly scalable	User-friendly
	AWS Comprehend	High	Highly scalable	User-friendly
	Edge AI (NVIDIA Jetson)	Very high (low latency)	Limited scalability	Requires expertise

8. Challenges and limitations

Despite advancements in sentiment analysis, several challenges persist due to the nature of social media data, computational constraints, linguistic diversity, and ethical considerations. Overcoming these issues is essential for enhancing the accuracy, reliability, and fairness of sentiment analysis models.

8.1. Data quality and noise issues

Social media data is inherently noisy, often containing informal language, abbreviations, emojis, and spelling errors. Additionally, spam, fake reviews, and bot-generated content can reduce data quality, leading to inaccurate sentiment classification. Incomplete or biased datasets further complicate the analysis. To address these challenges, effective preprocessing techniques such as text normalization, stopword removal, and outlier detection are essential for refining data before analysis.

8.2. Challenges in real-time sentiment analysis

Processing sentiment data in real time is computationally demanding due to the vast volume, rapid generation, and diverse nature of social media content. Ensuring low-latency processing while maintaining classification accuracy remains a significant challenge. Distributed computing frameworks such as Apache Spark and Flink are commonly used to manage real-time data streams, but issues like network latency and resource limitations persist. Moreover, detecting sarcasm, irony, and evolving slang in real-time analysis adds further complexity.

8.3. Gaps in multilingual and cross-cultural sentiment analysis

While machine learning (ML), deep learning (DL), and ensemble techniques have advanced the performance of sentiment analysis systems, most models still exhibit significant limitations when applied across multiple languages or cultural contexts. The dominance of English-language datasets has skewed model development and evaluation, leading to performance disparities when these models are deployed in multilingual or cross-cultural environments.

A key challenge lies in the representation of linguistic diversity. Most ML and DL models rely heavily on large annotated corpora, which are scarce for low-resource languages. This limitation hampers transferability and reduces the effectiveness of sentiment analysis for underrepresented populations. Even high-resource languages with distinct cultural sentiment expressions (e.g., Japanese, Arabic, or Hindi) may yield suboptimal results if models fail to capture region-specific semantics, idioms, or emotional cues.

Cross-cultural sentiment analysis is further complicated by cultural variance in emotional expression. For example, cultures differ in how directly emotions are expressed - Western cultures may favor explicit emotional statements, while many Asian cultures may express sentiment more subtly or indirectly. Models not attuned to these nuances can misclassify neutral or nuanced sentiment, especially when trained only on mono-cultural datasets.

Although multilingual models like mBERT and XLM-Roberta represent major progress, they are not immune to these limitations. Studies show performance degradation in languages with lower representation during pretraining, and these models often struggle with code-switching and contextual polysemy, both of which are prevalent in social media discourse.

There is a pressing need for research that addresses these gaps through culturally aware sentiment lexicons, context-sensitive embedding techniques, and collaborative dataset creation efforts spanning diverse linguistic communities. In addition, model evaluation should incorporate cross-cultural benchmarks to ensure generalizability and fairness across different user populations.

A more inclusive approach to multilingual and cross-cultural sentiment analysis will not only improve model performance but also support ethical AI practices by ensuring that sentiment technologies serve a globally diverse user base.

8.4. Complexities in multilingual sentiment analysis

Social media platforms host content in multiple languages, dialects, and mixed-language texts. Traditional sentiment analysis models, typically trained on English datasets, struggle with low-resource languages and cultural variations in sentiment expression. While translation tools can help, they often introduce errors that alter the intended meaning. Advanced multilingual models like mBERT and XLM-Roberta have improved cross-lingual sentiment analysis, but challenges remain in accurately processing code-switching and regional linguistic nuances [36], [71], [92].

Low-resource languages - those with limited annotated corpora, digital resources, and linguistic tools - pose a significant challenge to sentiment analysis systems. Many African, South Asian, and Indigenous languages fall into this category, often lacking the training data required to fine-tune or adapt general-purpose language models effectively. Consequently, sentiment expressed in such languages is frequently misclassified or ignored in large-scale analyses.

Recent advances have aimed to address this gap through techniques such as transfer learning, zero-shot learning, and the development of multilingual pre-trained language models. For example, XLM-Roberta has shown promising results in cross-lingual transfer by training on over 100 languages. However, its performance still varies significantly across languages depending on the volume and quality of the available training data [93], [94].

Efforts like the Masakhane project and IndicNLP have begun to create open-source datasets and benchmarks specifically tailored for African and Indian languages, respectively. These initiatives are crucial in fostering inclusive NLP research and ensuring that sentiment analysis tools are equitable across linguistic groups [95-97].

Despite these strides, fundamental limitations remain. Code-switching - where users alternate between languages within a single post - and culturally nuanced expressions continue to confound even state-of-the-art models. Moreover, regional idioms, sarcasm, and context-dependent sentiment expressions are often lost in translation or misinterpreted by general models.

To truly advance multilingual sentiment analysis, especially in underrepresented regions, there is a growing need for culturally aware models, community-driven dataset creation, and evaluation frameworks that reflect the linguistic diversity of global social media content.

8.5. Ethical and privacy concerns

The application of big data in sentiment analysis raises ethical and privacy-related concerns. Collecting and analyzing personal opinions from social media can compromise user privacy if the data is not anonymized. Regulatory frameworks such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) impose strict guidelines on data collection, processing, and usage. Additionally, biased algorithms may reinforce stereotypes, leading to unfair sentiment classification. Ensuring transparency, fairness, and adherence to data protection policies is crucial for ethical sentiment analysis.

With the rise of automated sentiment analysis systems, ethical AI policies are gaining prominence. Regulatory initiatives like the European Regulation on AI (EU AI Act) and frameworks aligned with GDPR and CCPA emphasize transparency, fairness, and accountability in algorithmic decision-making. Incorporating explainable AI (XAI) methods is vital to ensure that sentiment models align with these ethical standards and public expectations.

8.6. Interdisciplinary perspectives in sentiment analysis

While sentiment analysis is often approached from computational and data-driven standpoints, its effectiveness and depth are significantly enhanced through interdisciplinary integration. Insights from psychology and sociology provide foundational understanding of human emotions and social behavior, which can inform the development of more context-aware and culturally sensitive sentiment analysis models.
From a psychological perspective, emotional modeling plays a critical role in identifying and interpreting sentiment beyond surface-level keywords. Emotions are not binary but exist on a spectrum, with variations in intensity and polarity. Models rooted in affective computing and psychological theories such as the Circumplex Model of Affect [98] or Plutchik's Wheel of Emotions [99] offer richer sentiment categorization by incorporating dimensions like arousal, dominance, and valence. Integrating these models helps systems distinguish between emotions like anger and frustration or happiness and contentment, subtleties that purely lexical approaches may overlook.
Sociology contributes an understanding of how culture, community norms, and social identity shape the way people express emotions online. For example, expressions of grief, sarcasm, or approval may vary significantly across cultural or linguistic groups. Without this contextual awareness, sentiment models risk misclassification and bias. Sociolinguistic patterns - such as the use of irony in Western discourse or indirect speech in many Asian cultures - further illustrate the need for sociologically informed sentiment analysis frameworks [98-101].
Incorporating interdisciplinary methods also promotes ethical AI design by mitigating algorithmic bias and enhancing inclusivity. Psychological validation of sentiment categories, along with sociological input on language usage norms, can help build more accurate and fair sentiment analysis systems [102].

By addressing these challenges through advanced data processing techniques, robust machine learning models, and ethical data handling practices, sentiment analysis can become more accurate, inclusive, and responsible.
Given these persistent challenges-ranging from data quality issues to ethical complexities becomes essential to explore emerging technologies and novel methodologies. The next section delves into future directions that hold the potential to overcome current limitations and reshape sentiment analysis in a rapidly evolving digital landscape.
Addressing these limitations not only calls for robust technical interventions but also paves the way for strategic innovation. The subsequent section explores how emerging technologies and paradigms may tackle these ongoing challenges.

9. Future directions and opportunities

Building upon the identified limitations, this section outlines forward-looking strategies and technologies that hold potential to reshape sentiment analysis practices in alignment with evolving societal and computational needs.
With continuous advancements in technology, sentiment analysis is set to experience significant improvements. Innovations in artificial intelligence (AI), big data analytics, quantum computing, and real-time AI solutions will enhance its accuracy, scalability, and applicability. Additionally, expanding applications in various industries will create new opportunities for research and practical implementation.

9.1. Enhancements in AI and big data for sentiment analysis

The integration of AI and big data is revolutionizing sentiment analysis by enabling the efficient processing of vast amounts of unstructured social media data. Transformer-based deep learning models, such as BERT and GPT, continue to improve contextual understanding, leading to more precise sentiment classification. Automated Machine Learning (AutoML) and reinforcement learning are also being explored to streamline feature selection and enhance model adaptability. Moreover, federated learning techniques enable model training across multiple devices while maintaining data privacy, ensuring more secure and distributed sentiment analysis.

9.2. Impact of quantum computing and edge AI

Quantum computing has the potential to transform sentiment analysis by solving complex optimization and probabilistic modeling problems at unprecedented speeds. Quantum-enhanced machine learning algorithms, such as Quantum Support Vector Machines (QSVMs) and Quantum Boltzmann Machines, are being actively researched for their ability to accelerate training on big datasets [103]. For instance, the work by Rui Zhang et al. (2023) and Yao Chong Li et al. (2025) explores how quantum circuits can be integrated into hybrid neural networks to improve performance in classification tasks, which is central to sentiment analysis [104], [105]. Furthermore, IBM and Google have demonstrated quantum hardware prototypes capable of executing small-scale quantum ML tasks, laying the groundwork for future practical applications [106].
While fully realizing quantum sentiment analysis remains a long-term goal, preliminary research has shown promise. For example, few studies explored the use of quantum-enhanced feature spaces for natural language processing tasks, suggesting potential efficiency gains over classical methods. Such developments indicate a growing interest in applying quantum algorithms to NLP challenges, including emotion and sentiment recognition [107], [108].
Quantum computing has begun to show potential in accelerating machine learning operations through quantum-enhanced algorithms. Preliminary research in quantum natural language processing (QNLP), such as studies by Cambridge Quantum, suggests possibilities for improved data representation and processing efficiency. While still nascent, these developments indicate promising avenues for future sentiment analysis frameworks.
Quantum computing holds potential for accelerating complex NLP computations; however, its practical impact on sentiment analysis is still largely theoretical. Preliminary research suggests possible applications in optimization and language modeling, but further empirical studies are needed to validate these opportunities.
Meanwhile, Edge AI is witnessing rapid advancements and already demonstrates real-world impact. It enables real-time sentiment analysis on decentralized devices like smartphones, IoT sensors, and wearable technologies. This approach reduces latency, enhances data privacy, and minimizes dependency on cloud infrastructure [109]. Several commercial applications, such as emotion detection in mobile apps and real-time sentiment tracking in smart surveillance systems, have adopted Edge AI to provide low-latency responses [110]. A notable example is the deployment of Edge AI in customer interaction systems by companies like NVIDIA and Qualcomm, which leverage lightweight deep learning models for on-device sentiment classification [111], [112].

The convergence of quantum computing and Edge AI in the future could enable decentralized quantum-inference systems, though this remains speculative. However, ongoing research and pilot applications in both domains signal strong potential for their integration into scalable, intelligent sentiment analysis systems.

9.3. Advancements in real-time and explainable AI

The growing complexity of sentiment analysis models - especially deep learning and ensemble techniques - has intensified the need for interpretability. As a result, XAI frameworks are increasingly being explored to bridge the gap between model performance and transparency. Integrating such interpretable models is vital for trust, regulatory compliance, and informed decision-making, especially in high-stakes domains like healthcare and finance.

The demand for real-time sentiment analysis is increasing, particularly in industries like financial forecasting and crisis management. Streaming data architectures, such as Apache Kafka, are being utilized to process sentiment data instantaneously, allowing for quick decision-making. Additionally, XAI is gaining prominence by providing transparency in sentiment classification. Techniques like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) help users understand model predictions, addressing the challenges associated with deep learning's "black-box" nature and fostering trust in AI-driven sentiment analysis.

9.4. Expanding applications of sentiment analysis

Sentiment analysis has transitioned from a research concept to a practical tool that supports decision-making across industries. Two of the most impactful domains are business and healthcare, where sentiment insights are used to improve customer experience, inform strategic planning, and support mental health interventions [62], [113].

In the business sector, companies use sentiment analysis to monitor brand perception, track product reception, and evaluate customer satisfaction in real time [6], [114]. For instance, major e-commerce platforms like Amazon and Flipkart analyze customer reviews and social media mentions to assess product sentiment, enabling data-driven adjustments to inventory, marketing, or product design [115-117]. A practical example is Coca-Cola’s use of sentiment tracking tools during product launches to monitor audience reaction and adjust campaign messaging dynamically. Similarly, financial institutions analyze investor sentiment from news articles and social media feeds to forecast market trends and inform investment decisions [118].

In healthcare, sentiment analysis plays a critical role in understanding patient experiences and mental health trends. Hospitals and clinics mine patient feedback from online portals or surveys to identify dissatisfaction points in service delivery [119], [120]. More innovatively, mental health researchers have employed sentiment analysis on social media platforms such as Twitter and Reddit to detect early signs of depression, anxiety, or suicidal ideation [121]. One notable example is the use of sentiment-based monitoring tools during the COVID-19 pandemic to gauge public sentiment toward vaccines and health policies, which helped shape targeted communication strategies by public health agencies [122].

In business, sentiment analysis enables companies to evaluate product reception by analyzing customer reviews on platforms like Amazon or Twitter, guiding marketing strategies and product redesigns. In healthcare, patient sentiment on forums and health apps can help track mental well-being or detect early signs of depression, supporting personalized interventions and resource allocation.

These case studies highlight how sentiment analysis not only extracts useful patterns from vast amounts of unstructured data but also translates them into actionable insights. Integrating sentiment analysis into operational workflows empowers organizations to respond to emerging issues promptly, align with user expectations, and improve service outcomes.

Sentiment analysis is finding new applications across emerging fields. In the metaverse, it is being employed to assess user emotions in virtual reality (VR) environments, enhancing immersive digital experiences. Decentralized social media platforms are incorporating sentiment analysis to track trends while preserving user privacy. In the healthcare sector, it plays a crucial role in mental health monitoring by analyzing social media content for indicators of emotional distress. Additionally, smart cities are utilizing sentiment analysis to gauge public sentiment on policies, urban infrastructure, and transportation systems. These expanding applications highlight the growing influence and versatility of sentiment analysis in diverse industries.

Future work should explore hybrid models that fuse psychological theories of emotion with sociocultural context modeling. Collaboration between computational scientists, psychologists, and sociologists can foster the development of robust sentiment analysis frameworks that are both technically sound and human-centered.

As sentiment analysis continues to evolve, advancements in AI, computing power, and ethical AI practices will drive its adoption, making it more accurate, scalable, and impactful across multiple domains. Future research may investigate how federated learning frameworks can preserve data privacy while improving sentiment classification accuracy across decentralized social media platforms. Similarly, exploring the integration of QNLP could reveal new computational efficiencies in semantic analysis. Methodologies involving simulation-based modeling and privacy-aware architectures may help validate these concepts. These avenues present exciting opportunities for advancing scalable, secure, and interpretable sentiment analysis systems.

10. Conclusion

10.1. Summary of key findings

This review provided an in-depth analysis of sentiment analysis in social media using big data, discussing core concepts, methodologies, and technological advancements. Traditional machine learning and deep learning approaches, along with big data frameworks like Hadoop and Spark, have proven essential in extracting meaningful sentiment insights. Despite these advancements, several challenges persist, including data quality issues, real-time processing constraints, multilingual complexities, and ethical concerns. Additionally, the study explored emerging technologies like AI-driven innovations, quantum computing, and novel industrial applications as potential future directions.

10.2. Implications for researchers and industry professionals

For researchers, this study underscores critical areas for further investigation, including enhancing sentiment analysis accuracy in underrepresented languages and developing more interpretable AI models. The integration of real-time AI and edge computing presents new research opportunities to improve processing efficiency. For industry professionals, harnessing big data sentiment analysis can enhance decision-making across sectors such as business, healthcare, and manufacturing. Ensuring ethical AI deployment and addressing privacy concerns will be crucial in promoting responsible sentiment analysis practices.
This review underscores not only the technical advancements in sentiment analysis but also the importance of transparency and interpretability in model design. Addressing current gaps, such as the development of explainable and multilingual models, is essential for enhancing real-world applicability and fostering user trust.

10.3. Final thoughts and recommendations

Sentiment analysis in social media, driven by big data, remains a dynamic and evolving field with significant potential. Future research should focus on overcoming existing challenges while incorporating advanced technologies such as XAI and federated learning. Collaboration among policymakers, industry leaders, and researchers is essential to establish ethical guidelines for responsible sentiment analysis. By refining AI-driven sentiment analysis techniques, organizations can gain deeper insights into public opinion, leading to more informed and strategic decision-making.

References

[1] Margarita Rodríguez-Ibanez, Antonio Casánez-Ventura, Félix Castejón-Mateos, and Pedro-Manuel Cuenca-Jiménez, ‘A review on sentiment analysis from social media platforms’, *Expert systems with applications*, Vol. 223, Aug. 2023, pp. 119862, https://doi.org/10.1016/j.eswa.2023.119862.

[2] Nunik Afriliana, Ni Made Satvika Iswari, and Suryasari ‘Sentiment Analysis of User-Generated Content: A Bibliometric Analysis’, *Journal of System and Management Sciences*, Vol. 12, No. 6, Dec. 2022, pp. 583-598, https://doi.org/10.33168/JSMS.2022.0634.

[3] A. Shaji George, and T. Baskar, ‘Leveraging Big Data and Sentiment Analysis for Actionable Insights: A Review of Data Mining Approaches for Social Media’, *Partners Universal International Innovation Journal (PUIIJ)*, Vol. 2, No. 4, Aug. 2024, pp. 39-59.

[4] Alanoud Alotaibi, and Farrukh Nadeem, ‘Leveraging Social Media and Deep Learning for Sentiment Analysis for Smart Governance: A Case Study of Public Reactions to Educational Reforms in Saudi Arabia’, *Computers*, Vol. 13, No. 11, Oct. 2024, pp. 280 https://doi.org/10.3390/computers13110280.

[5] Md Abdullah Al Montaser, Bishnu Padh Ghosh, Ayan Barua, Fazle Karim, Bimol Chandra Das, Reza E Rabbi Shawon, and Muhammad Shoyaibur Rahman Chowdhury, ‘Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA’, *Edelweiss Applied Science and Technology*, Vol. 9, No. 1, Jan. 2025, pp. 515-535, https://doi.org/10.55214/25768484.v9i1.4164.

[6] Anant Bhardwaj, Pradyumn Pratap Singh, Astha Bharti, Apurva Singh Parihar, and Sandeep Kaur, ‘Social Media Sentiment Analysis for Brand Monitoring’, *International Journal of Engineering Research & Technology (IJERT)*, Vol. 13, Issue 10, Oct. 2024. Available at: https://www.ijert.org/research/social-media-sentiment-analysis-for-brand-monitoring-IJERTV13IS100134.pdf.

[7] Joyce Y. M. Nip, and Benoit Berthelier, ‘Social Media Sentiment Analysis’, *Encyclopedia*, Vol. 4, No. 4, Oct. 2024, pp. 1590-1598, , https://doi.org/10.3390/encyclopedia4040104.

[8] Qianwen Ariel Xu, Victor Chang, and Chrisina Jayne, ‘A systematic review of social media-based sentiment analysis: Emerging trends and challenges’, *Decision Analytics Journal*, vol. 3, p. 100073, Jun. 2022, https://doi.org/10.1016/j.dajour.2022.100073.

[9] Erfan Bakhtiari Ramezani, ‘Sentiment analysis applications using deep learning advancements in social networks: A systematic review’, *Neurocomputing*, Vol. 634, Jun. 2025, pp. 129862, https://doi.org/10.1016/j.neucom.2025.129862.

[10] Yanying Mao, Qun Liu, and Yu Zhang, ‘Sentiment analysis methods, applications, and challenges: A systematic literature review’, *Journal of King Saud University - Computer and Information Sciences*, Vol. 36, No. 4, Apr. 2024, pp. 102048, https://doi.org/10.1016/j.jksuci.2024.102048.

[11] Maite Taboada, ‘Sentiment Analysis: An Overview from Linguistics’, *Annual Review of Linguistics*, Vol. 2, No. 1, Jan. 2016, pp. 325-347, https://doi.org/10.1146/annurev-linguistics-011415-040518.

[12] Mohammed H. Abd El-Jawad, Rania Hodhod, and Yasser M. K. Omar, ‘Sentiment Analysis of Social Media Networks Using Machine Learning’, in *2018 14th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt: IEEE, Dec. 2018, pp. 174–176, https://doi.org/10.1109/ICENCO.2018.8636124.

[13] Julian Villodre, ‘Implementing social media monitoring in the public sector: a four-model approach’. In Sarah Giest, Bram Klievink, Alex Ingrams, and Matthew M. Young (eds.) *Handbook on Governance and Data Science*, Edward Elgar, 2025. https://doi.org/10.4337/9781035301348.00014.

[14] Xieling Chen, Haoran Xie, Zongxi Li, Han Zhang, Xiaohui Tao, and Fu Lee Wang, ‘Sentiment analysis for stock market research: A bibliometric study’, *Natural Language Processing Journal*, Vol. 10, Mar. 2025, pp. 100125, https://doi.org/10.1016/j.nlp.2025.100125.

[15] Pragati Mishra, Chetan Agrawal, and Divya Envey, ‘Advancements and Challenges in Sentiment Analysis: A Comprehensive Review of Frameworks and Future Trends’, *International Journal of Engineering Applied Science and Management (IJEASM)*, Vol. 4, Issue 12, December 2023, pp. 1-8, Paper ID: 2023/IJEASM/4/2023/1932a, Available at: https://www.ijeasm.com/PublishedPaper/4Vol/Issue12/2023IJEASM420231932a-4d089232-3d34-40ed-8b58-11c961c2cb8f48503.pdf.

[16] Penubaka Balaji, O. Nagaraju, and D. Haritha, ‘Levels of sentiment analysis and its challenges: A literature review’, in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, Andhra Pradesh, India: IEEE, Mar. 2017, pp. 436–439, https://doi.org/10.1109/ICBDACI.2017.8070879.

[17] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni, ‘A survey on sentiment analysis methods, applications, and challenges’, *Artificial Intelligence Review*, Volume 55, (2022), pp. 5731-5780, https://doi.org/10.1007/s10462-022-10144-1.

[18] Mahander Kumar, Lal Khan, and Hsien-Tsung Chang, ‘Evolving techniques in sentiment analysis: a comprehensive review’, *PeerJ Computer Science*, 11:e2592, January 2025, https://doi.org/10.7717/peerj-cs.2592.

[19] Walaa Medhat, Ahmed Hassan, and Hoda Korashy, ‘Sentiment analysis algorithms and applications: A survey’, *Ain Shams Engineering Journal*, Vol. 5, No. 4, Dec. 2014, pp. 1093-1113, https://doi.org/10.1016/j.asej.2014.04.011.

[20] Egil Rønningstad, Erik Velldal, and Lilja Øvrelid, ‘Entity-Level Sentiment Analysis (ELSA): An Exploratory Task Survey’, in *Proceedings of the 29th International Conference on Computational Linguistics*, N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, and S.-H. Na, Eds., Gyeongju, Republic of Korea: International Committee on Computational Linguistics, Oct. 2022, pp. 6773–6783. Available at: https://aclanthology.org/2022.coling-1.589/.

[21] Sandeep Singh, and Harjot Kaur, ‘Comparative Sentiment Analysis Through Traditional and Machine Learning-Based Approach’, in *Cognitive Behavior and Human Computer Interaction Based on Machine Learning Algorithm*, 1st ed., S. Kumar, R. Raja, S. Tiwari, and S. Rani, Eds., Wiley, 2021, pp. 315–338. https://doi.org/10.1002/9781119792109.ch14.

[22] Ajeet Ram Pathak, Manjusha Pandey, and Siddharth Rautaray, ‘Topic-level sentiment analysis of social media data using deep learning’, *Applied Soft Computing*, Vol. 108, Sep. 2021, pp. 107440, https://doi.org/10.1016/j.asoc.2021.107440.

[23] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede, ‘Lexicon-Based Methods for Sentiment Analysis’, *Computational Linguistics*, Vol. 37, No. 2, Jun. 2011, pp. 267-307, https://doi.org/10.1162/COLI_a_00049.

[24] Taminul Islam, Md. Alif Sheakh, Md. Rezwane Sadik, Mst. Sazia Tahosin, Md. Musfiqur Rahman Foysal, Jannatul Ferdush, and Mahbuba Begum ‘Lexicon and Deep Learning-Based Approaches in Sentiment Analysis on Short Texts’, *Journal of Computer and Communications*, Vol. 12, No. 01, 2024, pp. 11-34, https://doi.org/10.4236/jcc.2024.121002.

[25] Hesham Allam, Lisa Makubvure, Benjamin Gyamfi, Kwadwo Nyarko Graham and Kehinde Akinwolere, ‘Text Classification: How Machine Learning Is Revolutionizing Text Categorization’, *Information*, Vol. 16, No. 2, Feb. 2025, pp. 130, https://doi.org/10.3390/info16020130.

[26] Nikhil Sanjay Suryawanshi, ‘Sentiment analysis with machine learning and deep learning: A survey of techniques and applications’, *International Journal of Science and Research Archive*, Vol. 12, No. 2, Jul. 2024, pp. 005–015, https://doi.org/10.30574/ijsra.2024.12.2.1205.

[27] Zhenyu Li, Zongfeng Zou, ‘Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform’, *Journal of King Saud University - Computer and Information Sciences*, Vol. 36, No. 3, Mar. 2024, pp. 102010, https://doi.org/10.1016/j.jksuci.2024.102010.

[28] Mahdi Hajiali, ‘Big data and sentiment analysis: A comprehensive and systematic literature review’, *Concurrency and Computation*, Vol. 32, No. 14, Jul. 2020, pp. e5671, https://doi.org/10.1002/cpe.5671.

[29] Ajeet Ram Pathak, Manjusha Pandey, and Siddharth Rautaray, ‘Topic-level sentiment analysis of social media data using deep learning’, *Applied Soft Computing*, Vol. 108, Sep. 2021, pp. 107440, https://doi.org/10.1016/j.asoc.2021.107440.

[30] Parneet Kaur, ‘Sentiment analysis using web scraping for live news data with machine learning algorithms’, *Materials Today: Proceedings*, Vol. 65, Part 8, 2022, pp. 3333–3341, https://doi.org/10.1016/j.matpr.2022.05.409.

[31] Iryna Mysiuk, and Roman Shuvar, ‘Automating Web Scraping of User Comments for Sentiment Analysis in Social Networks’, in *2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)*, Lviv, Ukraine: IEEE, Sep. 2023, pp. 77–81, https://doi.org/10.1109/ELIT61488.2023.10310867.

[32] Xiaoqin Shelley Zhang, Xiaojie Zhang, and Pavan Kaparthi, ‘Combat Information Overload Problem in Social Networks With Intelligent Information-Sharing and Response Mechanisms’, *IEEE Transactions on Computational Social Systems*, Vol. 7, No. 4, Aug. 2020, pp. 924–939, https://doi.org/10.1109/TCSS.2020.3001093.

[33] Sampsa Suvivuo, and Virpi Tuunainen, ‘Challenges and Solutions in Qualitative Big Data Research: A Methodological Literature Review’, *Communications of the Association for Information Systems*, Vol. 55, No. 1, 2024, pp. 37–76, https://doi.org/10.17705/1CAIS.05502.

[34] Ravindra Kumar Singh, and Harsh Kumar Verma, ‘Effective Parallel Processing Social Media Analytics Framework’, *Journal of King Saud University - Computer and Information Sciences*, Vol. 34, No. 6, Jun. 2022, pp. 2860–2870, https://doi.org/10.1016/j.jksuci.2020.04.019.

[35] Ahmad M. Alghamdi, Salvatore Flavio Pileggi, and Osama Sohaib, ‘Social Media Analysis to Enhance Sustainable Knowledge Management: A Concise Literature Review’, *Sustainability*, Vol. 15, No. 13, Jun. 2023, pp. 9957, https://doi.org/10.3390/su15139957.

[36] Koena Ronny Mabokela, Turgay Celik, and Mpho Raborife, ‘Multilingual Sentiment Analysis for Under-Resourced Languages: A Systematic Review of the Landscape’, *IEEE Access*, Vol. 11, 2023, pp. 15996–16020, https://doi.org/10.1109/ACCESS.2022.3224136.

[37] Michael James, ‘The Ethical and Legal Implications of Using Big Data and Artificial Intelligence for Public Relations Campaigns in the United States’, *International Journal of Communication and Public Relation*, Vol. 9, No. 1, Jan. 2024, pp. 38–52, https://doi.org/10.47604/ijcpr.2273.

[38] Louis Owusu-Berko, ‘Harnessing Big Data, Machine Learning, and Sentiment Analysis to Optimize Customer Engagement, Loyalty, and Market Positioning’, *International Journal of Computer Applications Technology and Research*, Vol. 14, No. 3, Feb. 2025, pp. 1-16, https://doi.org/10.7753/IJCATR1403.1001.

[39] Saurav Pradha, Malka N. Halgamuge, and Nguyen Tran Quoc Vinh, ‘Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data’, in *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, Da Nang, Vietnam: IEEE, Oct. 2019, pp. 1–8. https://doi.org/10.1109/KSE.2019.8919368.

[40] Raed Abdulkareem Hasan, Tole Sutikno, and Mohd Arfian Ismail, ‘A Review on Big Data Sentiment Analysis Techniques’, *Mesopotamian Journal of Big Data*, Vol. 2021, Jan. 2021, pp. 6–13, https://doi.org/10.58496/MJBD/2021/002.

[41] Rajesh Kumar Das, Mirajul Islam, Md Mahmudul Hasan, Sultana Razia, Mocksidul Hassan, and Sharun Akter Khushbu, ‘Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models’, *Heliyon*, Vol. 9, No. 9, Sep. 2023, p. e20281, https://doi.org/10.1016/j.heliyon.2023.e20281.

[42] Nabaa Muhammad Diaa, Saba Sabah Ahmed. Hayder Mahmood Salman, and Wafaa Adnan Sajid, ‘Statistical Challenges in Social Media Data Analysis Sentiment Tracking and Beyond’, *Journal of Ecohumanism*, Vol. 3, No. 5, Sep. 2024, pp. 365–384, https://doi.org/10.62754/joe.v3i5.3912.

[43] Mahander Kumar, Lal Khan, and Hsien-Tsung Chang, ‘Evolving techniques in sentiment analysis: a comprehensive review’, *PeerJ Computer Science*, Vol. 11, Jan. 2025, p. e2592, https://doi.org/10.7717/peerj-cs.2592.

[44] Aish Albladi, Minarul Islam, and Cheryl Seals, ‘Sentiment Analysis of Twitter Data Using NLP Models: A Comprehensive Review’, *IEEE Access*, Vol. 13, Feb. 2025, pp. 30444–30468, https://doi.org/10.1109/ACCESS.2025.3541494.

[45] Jamin Rahman Jim, Md Apon Riaz Talukder, Partha Malakar, Md Mohsin Kabir, Kamruddin Nur, and M.F. Mridha, ‘Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review’, *Natural Language Processing Journal*, Vol. 6, Mar. 2024, p. 100059, https://doi.org/10.1016/j.nlp.2024.100059.

[46] Najeeb Abdulazez Alabdulkarim, Mohd Anul Haq, and Jayadev Gyani, ‘Exploring Sentiment Analysis on Social Media Texts’, *Engineering, Technology & Applied Science Research*, vol. 14, No. 3, Jun. 2024, pp. 14442–14450, https://doi.org/10.48084/etasr.7238.

[47] Kian Long Tan, Chin Poo Lee, and Kian Ming Lim, ‘A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research’, *Applied Sciences*, vol. 13, No. 7, Apr. 2023, p. 4550, https://doi.org/10.3390/app13074550.

[48] Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, Mounika Bagadi, and Mounika Belusonti, ‘Text based Sentiment Analysis using LSTM’, *International Journal of Engineering Research & Technology (IJERT)*, Vol. 9, No. 5, May 2020, p. IJERTV9IS050290, https://doi.org/10.17577/IJERTV9IS050290.

[49] Kian Long Tan, Chin Poo Lee, Kian Ming Lim, and Kalaiarasi Sonai Muthu Anbananthen, ‘Sentiment Analysis with Ensemble Hybrid Deep Learning Model’, *IEEE Access*, Vol. 10, Sept. 2022, pp. 103694–103704, https://doi.org/10.1109/ACCESS.2022.3210182.

[50] Aman Sawarn, Ankit, and Monika Gupta, ‘Comparative Analysis of Bagging and Boosting Algorithms for Sentiment Analysis’, *Procedia Computer Science*, Vol. 173, 2020, pp. 210–215, https://doi.org/10.1016/j.procs.2020.06.025.

[51] Dimple Tiwari, Bharti Nagpal, Bhoopesh Singh Bhati, Ashutosh Mishra, and Manoj Kumar, ‘A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques’, *Artificial Intelligence Review*, Vol. 56, No. 11, Nov. 2023, pp. 13407–13461, https://doi.org/10.1007/s10462-023-10472-w.

[52] Ahmed Alsayat, ‘Improving sentiment analysis for social media applications using an ensemble deep learning language model’, *Arabian Journal for Science and Engineering*, Vol. 47, No. 2, Feb. 2022, pp. 2499–2511, https://doi.org/10.1007/s13369-021-06227-w.

[53] Su Cui, Yiliang Han, Yifei Duan, Yu Li, Shuaishuai Zhu, and Chaoyue Song, ‘A two-stage voting-boosting technique for ensemble learning in social network sentiment classification’, *Entropy*, Vol. 25, No. 4, Mar. 2023, p. 555, https://doi.org/10.3390/e25040555.

[54] Kottala Sri Yogi, Dankan Gowda V., Galiveeti Poornima, Kdv Prasad, Srinivas. D., P. Vishnu Prasanth, 'Comparative analysis of machine learning techniques for detecting sentiments in social media', *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, Namakkal, India, 15-16 March 2024, pp. 1-6, May 2024, https://doi.org/10.1109/AIMLA59606.2024.10531524.

[55] Kruttika Jain, and Shivani Kaushal, 'A comparative study of machine learning and deep learning techniques for sentiment analysis', *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 29-31 August 2018, pp. 483-487, July 2019, https://doi.org/10.1109/ICRITO.2018.8748793.

[56] Prajval Sudhir, and Varun Deshakulkarni Suresh, ‘Comparative study of various approaches, applications and classifiers for sentiment analysis’, *Global Transitions Proceedings*, Vol. 2, No. 2, Nov. 2021, pp. 205-211, https://doi.org/10.1016/j.gltp.2021.08.004.

[57] Talha Ahmed Khan, Rehan Sadiq, Zeeshan Shahid, Muhammad Mansoor Alam, and Mazliham Bin Mohd Su'ud, ‘Sentiment analysis using support vector machine and random forest’, *Journal of Informatics and Web Engineering*, Vol. 3, No. 1, Feb. 2024, pp. 67–75, https://doi.org/10.33093/jiwe.2024.3.1.5.

[58] Abinash Tripathy, Ankit Agrawal, and Santanu Kumar Rath, ‘Classification of sentiment reviews using n-gram machine learning approach’, *Expert Systems with Applications*, Vol. 57, Sep. 2016, pp. 117–126, https://doi.org/10.1016/j.eswa.2016.03.028.

[59] Noura A Semary, Wesam Ahmed, Khalid Amin, Paweł Pławiak, and Mohamed Hammad, ‘Enhancing machine learning-based sentiment analysis through feature extraction techniques’, *PLoS ONE*, Vol. 19, No. 2, Feb. 2024, p. e0294968, https://doi.org/10.1371/journal.pone.0294968.

[60] Jochen Hartmann, Mark Heitmann, Christian Siebert, and Christina Schamp, ‘More than a feeling: accuracy and application of sentiment analysis’, *International Journal of Research in Marketing*, Vol. 40, No. 1, Mar. 2023, pp. 75–87, https://doi.org/10.1016/j.ijresmar.2022.05.005.

[61] Mrunmayee Bachate, and Suchitra S., ‘Sentiment analysis and emotion recognition in social media: A comprehensive survey’, *Applied Soft Computing*, Vol. 174, Apr. 2025, p. 112958, https://doi.org/10.1016/j.asoc.2025.112958.

[62] Neeraj Anand Sharma, A. B. M. Shawkat Ali, and Muhammad Ashad Kabir, ‘A review of sentiment analysis: tasks, applications, and deep learning techniques’, *International Journal of Data Science and Analytics*, Vol. 19, No. 3, Apr. 2025, pp. 351–388, https://doi.org/10.1007/s41060-024-00594-x.

[63] Sayyida Tabinda Kokab, Sohail Asghar, and Shehneela Naz, ‘Transformer-based deep learning models for the sentiment analysis of social media data’, *Array*, Vol. 14, Jul. 2022, p. 100157, https://doi.org/10.1016/j.array.2022.100157.

[64] Kamran Razzaq, and Mahmood Shah, ‘Machine learning and deep learning paradigms: from techniques to practical applications and research frontiers’, *Computers*, Vol. 14, No. 3, Mar. 2025, p. 93, https://doi.org/10.3390/computers14030093.

[65] Babacar Gaye, Dezheng Zhang, and Aziguli Wulamu, ‘A tweet sentiment classification approach using a hybrid stacked ensemble technique’, *Information*, Vol. 12, No. 9, Sep. 2021, p. 374, https://doi.org/10.3390/info12090374.

[66] Petr Hajek, and Michal Munk, ‘Speech emotion recognition and text sentiment analysis for financial distress prediction’, *Neural Computing and Applications*, Vol. 35, No. 29, Oct. 2023, pp. 21463-21477, https://doi.org/10.1007/s00521-023-08470-8.

[67] Shaymaa E. Sorour, Abdulrahman Alojail, Amr El-Shora, Ahmed E. Amin, and Amr A. Abohany, ‘A Hybrid Deep Learning Approach for Enhanced Sentiment Classification and Consistency Analysis in Customer Reviews’, *Mathematics*, Vol. 12, No. 23, Dec. 2024, p. 3856, https://doi.org/10.3390/math12233856.

[68] Jacqueline Kazmaier, and Jan H. van Vuuren, "The power of ensemble learning in sentiment analysis", *Expert Systems with Applications*, Volume 187, January 2022, 115819, https://doi.org/10.1016/j.eswa.2021.115819.

[69] M. Khairul Anam, Tri Putri Lestari, Helda Yenni, Torkis Nasution, and Muhammad Bambang Firdaus, ‘Enhancement of Machine Learning Algorithm in Fine-grained Sentiment Analysis Using the Ensemble’, *ECTI Transactions on Computer and Information Technology*, Volume 19, No. 2, Mar. 2025, pp. 159–167, https://doi.org/10.37936/ecti-cit.2025192.257815.

[70] Jyoti Yadav, ‘Sentiment Analysis on Social Media’, *Qeios*, Jan. 09, 2023. https://doi.org/10.32388/YF9X04.

[71] Yusuf Aliyu, Aliza Sarlan, Kamaluddeen Usman Danyaro, Abdullahi Sani B. A. Rahman, and Mujaheed Abdullahi, ‘Sentiment Analysis in Low-Resource Settings: A Comprehensive Review of Approaches, Languages, and Data Sources’, *IEEE Access*, vol. 12, pp. 66883–66909, 2024, https://doi.org/10.1109/ACCESS.2024.3398635.

[72] Hemn Barzan Abdalla, and Yulia Kumar, Yue Zhao and Davide Tosi, ‘A Comprehensive Survey of MapReduce Models for Processing Big Data’, *Big Data and Cognitive Computing*, vol. 9, no. 4, 77, March 2025. https://doi.org/10.3390/bdcc9040077.

[73] K. Sowmya, K. Chandu, and N. HarshaVardhan Reddy, ‘Sentiment Analysis Using Hadoop Framework and Deep Learning Algorithm’, *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, vol. 13, issue 4, April 2024, pp. 5154 - 5161. Available at: https://www.ijirset.com/upload/2024/april/295_Sentiment.pdf.

[74] Deniz Kilinç, ‘A spark-based big data analysis framework for real-time sentiment prediction on streaming data’, *Software: Practice and Experience*, vol. 49, no. 9, pp. 1352–1364, Sep. 2019, https://doi.org/10.1002/spe.2724.

[75] Uğur Kekevi , and Ahmet Arif Aydın, ‘Real-Time Big Data Processing and Analytics: Concepts, Technologies, and Domains’, *Computer Science*, Volume 7, Issue 2, Nov. 2022, pp. 111-123, https://doi.org/10.53070/bbd.1204112.

[76] Sudesh Rani, ‘Comparative Analysis of Big Data Analysis Tools: A Review of Features, Performance, and Applications’, *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 10, issue 6, June 2023, pp. 327 - 333. Available at: https://www.jetir.org/papers/JETIR2306837.pdf.

[77] Leonidas Theodorakopoulos, Alexandra Theodoropoulou and Yannis Stamatiou, ‘A State-of-the-Art Review in Big Data Management Engineering: Real-Life Case Studies, Challenges, and Future Research Directions’, *Eng*, vol. 5, no. 3, pp. 1266–1297, Jul. 2024, https://doi.org/10.3390/eng5030068.

[78] Youngsub Han, and Kwangmi Ko Kim, ‘Sentiment analysis on social media using morphological sentence pattern model’, in 2017 *IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, London, United Kingdom: IEEE, Jun. 2017, pp. 79–84. https://doi.org/10.1109/SERA.2017.7965710.

[79] Dipanjan Sarkar, ‘Sentiment Analysis’, in *Text Analytics with Python*, Berkeley, CA: Apress, 2019, pp. 567-629. https://doi.org/10.1007/978-1-4842-4354-1_9.

[80] Abdulrahman Radaideh, and Fikri Dweiri, ‘Sentiment Analysis Predictions in Digital Media Content using NLP Techniques’, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 11, 2023, https://doi.org/10.14569/IJACSA.2023.0141128.

[81] Dhara Ashish Darji, and Sachinkumar Anandpal Goswami, ‘The Comparative study of Python Libraries for Natural Language Processing (NLP)’, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, vol. 10, no. 2, pp. 499–512, Mar. 2024, https://doi.org/10.32628/CSEIT2410242.

[82] Muhammad Iqbal Faturohman, and Miftahol Arifin, ‘TextBlob-Based Sentiment Analysis of Tabungan Perumahan Rakyat (TAPERA) Policy: A Public Perception Study’, *J@ti Undip: Jurnal Teknik Industri*, vol. 20, no. 1, pp. 11–20, Feb. 2025, https://doi.org/10.14710/jati.20.1.11-20.

[83] Monali Bordoloi, and Saroj Kumar Biswas, ‘Sentiment analysis: A survey on design framework, applications and future scopes’, *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12505–12560, Nov. 2023, https://doi.org/10.1007/s10462-023-10442-2.

[84] Andrea Bonetti, Marcelino Martínez-Sober, Julio C. Torres, Jose M. Vega, Sebastien Pellerin, and Joan Vila-Francés, ‘Comparison between Machine Learning and Deep Learning Approaches for the Detection of Toxic Comments on Social Networks’, *Applied Sciences*, vol. 13, no. 10, p. 6038, May 2023, https://doi.org/10.3390/app13106038.

[85] Logan Ashbaugh, and Yan Zhang, ‘A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning’, *Computers*, vol. 13, no. 12, p. 340, Dec. 2024, https://doi.org/10.3390/computers13120340

[86] Mahander Kumar, Lal Khan, and Hsien-Tsung Chang, ‘Evolving techniques in sentiment analysis: a comprehensive review’, *PeerJ Computer Science*, vol. 11, p. e2592, Jan. 2025, https://doi.org/10.7717/peerj-cs.2592.

[87] Jiangao Deng, and Yue Liu, ‘Research on Sentiment Analysis of Online Public Opinion Based on RoBERTa–BiLSTM–Attention Model’, *Applied Sciences*, vol. 15, no. 4, p. 2148, Feb. 2025, https://doi.org/10.3390/app15042148.

[88] Louis Owusu-Berko, ‘Harnessing Big Data, Machine Learning, and Sentiment Analysis to Optimize Customer Engagement, Loyalty, and Market Positioning’, *International Journal of Computer Applications Technology and Research*, Feb. 2025.

[89] Sergiu C. Ivan, Robert Ş. Győrödi, and Cornelia A. Győrödi, ‘Sentiment Analysis Using Amazon Web Services and Microsoft Azure’, *Big Data and Cognitive Computing*, vol. 8, no. 12, p. 166, Nov. 2024, https://doi.org/10.3390/bdcc8120166.

[90] Thanveer Shaik, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, and Linda Galligan, ‘Sentiment analysis and opinion mining on educational data: A survey’, *Natural Language Processing Journal*, vol. 2, p. 100003, Mar. 2023, https://doi.org/10.1016/j.nlp.2022.100003.

[91] Prema Nedungadi, G. Veena, Kai-Yu Tang, Remya R. K. Menon, and Raghu Raman, 'AI Techniques and Applications for Online Social Networks and Media: Insights From BERTopic Modeling', *IEEE Access*, vol. 13, pp. 37389–37407, 2025, https://doi.org/10.1109/ACCESS.2025.3543795.
[92] Yusuf Aliyu, Aliza Sarlan, Kamaluddeen Usman Danyaro, Abdullahi Sani abd Rahman, Aminu Aminu Muazu, and Mustapha Yusuf Abubakar, 'Deep learning techniques for sentiment analysis in code-switched Hausa-English tweets', *International Journal of Information Management Data Insights*, vol. 5, no. 1, article no. 100330, Jun. 2025, https://doi.org/10.1016/j.jjimei.2025.100330.
[93] Muhamet Kastrati, Ali Shariq Imran, Ehtesham Hashmi, Zenun Kastrati, Sher Muhammad Daudpota, and Marenglen Biba, 'Unlocking language barriers: Assessing pre-trained large language models across multilingual tasks and unveiling the black box with Explainable Artificial Intelligence', *Engineering Applications of Artificial Intelligence*, vol. 149, article no. 110136, Jun. 2025, https://doi.org/10.1016/j.engappai.2025.110136.
[94] Kai Feng, Lan Huang, Kangping Wang, Wei Wei, and Rui Zhang, 'Prompt-based learning framework for zero-shot cross-lingual text classification', *Engineering Applications of Artificial Intelligence*, Vol. 133, Part E, article no. 108481, Jul. 2024, https://doi.org/10.1016/j.engappai.2024.108481.
[95] Anoop Kunchukuttan, The IndicNLP Library, 2020, https://github.com/anoopkunchukuttan/indic_nlp_ library/blob/master/docs/indicnlp. pdf.
[96] Brijeshkumar Y. Panchal and Apurva Shah, 'NLP Research: A Historical Survey and Current Trends in Global, Indic, and Gujarati Languages', *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, Gobichettipalayam, India, 2024, pp. 1263-1272, https://doi.org/10.1109/ICUIS64676.2024.10866047.
[97] David Ifeoluwa Adelani, Marek Masiak, et al., 'MasakhaNEWS: News Topic Classification for African languages', *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics* (Volume 1: Long Papers), pages 144–159 November 1–4, 2023. Association for Computational Linguistics. Available at: https://aclanthology.org/2023.ijcnlp-main.10.pdf.
[98] Alejandro De León Languré, and Mahdi Zareei, 'Evaluating the Effect of Emotion Models on the Generalizability of Text Emotion Detection Systems', *IEEE Access*, vol. 12, May 2024, https://doi.org/10.1109/ACCESS.2024.3401203.
[99] Alfonso Semeraro, Salvatore Vilella, and Giancarlo Ruffo, 'PyPlutchik: Visualising and comparing emotion-annotated corpora', *PLoS ONE*, vol. 16, no. 9, article no. e0256503, Sept. 2021, https://doi.org/10.1371/journal.pone.0256503.
[100] Diego Resende Faria, Abraham Itzhak Weinberg, and Pedro Paulo Ayrosa, 'Multimodal Affective Communication Analysis: Fusing Speech Emotion and Text Sentiment Using Machine Learning', *Applied Sciences*, vol. 14, no. 15, article no. 6631, Jul. 2024, https://doi.org/10.3390/app14156631.
[101] Fa Zhang, Jian Chen, Qian Tang, and Yan Tian, 'Evaluation of emotion classification schemes in social media text: an annotation-based approach', *BMC Psychology*, Volume 12, article no. 503, Sept. 2024, https://doi.org/10.1186/s40359-024-02008-w.
[102] Saif M. Mohammad, 'Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis', *Computational Linguistics*, vol. 48, no. 2, pp. 239–278, Jun. 2022, https://doi.org/10.1162/coli_a_00433.
[103] Giuseppe Buonaiuto, Raffaele Guarasci, Aniello Minutolo, Giuseppe De Pietro, and Massimo Esposito, 'Quantum transfer learning for acceptability judgements', *Quantum Machine Intelligence*, vol. 6, article no. 13, Mar 2024. https://doi.org/10.1007/s42484-024-00141-8.
[104] Rui Zhang, Jian Wang, Nan Jiang, and Zichen Wang, 'Quantum support vector machine without iteration', *Information Sciences*, vol. 635, pp. 25-41, Jul. 2023, https://doi.org/10.1016/j.ins.2023.03.106.
[105] YaoChong Li, Yi Qu, Ri-Gui Zhou, and Jing Zhang, 'QMLSC: A quantum multimodal learning model for sentiment classification', *Information Fusion,* vol. 120, article no. 103049, Aug 2025, https://doi.org/10.1016/j.inffus.2025.103049.
[106] Spyridon A. Pappas, 'Quantum Programming and its Strategic Applications in Artificial Intelligence Systems', *Ph. D. Thesis*, Apr 2025, Available at: https://science.edu.gr/research/sapappas/Quantum%20Programming%20AI.pdf.
[107] Kaduri Sreelakshmi, Vishal V. Rathi, K. Shanthi, Shikalgar Niyaj Dilavar, S. Logeswaran, and B. Subhi, 'Real-World Applications of Quantum-Enhanced Machine Learning Solutions', in *Advances in Computational Intelligence and Robotics*, Minu R.I., Nagarajan G., Martin Margala, Siva Shankar S., and Logashanmugam E., Eds., IGI Global, Apr 2025, pp. 81-106. https://doi.org/10.4018/979-8-3693-6225-9.ch004.
[108] Abu Kaisar Mohammad Masum, Anshul Maurya, Dhruthi Sridhar Murthy, Pratibha, and Naveed Mahmud, 'Hybrid Quantum-Classical Machine Learning for Sentiment Analysis', arXiv: arXiv:2310.10672, Oct 2023, https://doi.org/10.1109/ICMLA58977.2023.00149.
[109] Raghubir Singh, and Sukhpal Singh Gill, 'Edge AI: A survey', *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 71-92, Mar 2023, https://doi.org/10.1016/j.iotcps.2023.02.004.
[110] Elarbi Badidi, Karima Moumane, and Firdaous El Ghazi, 'Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review', *IEEE Access*, vol. 11, pp. 80543-80572, Aug 2023, https://doi.org/10.1109/ACCESS.2023.3300658
[111] Yiwen Xu, Tariq M. Khan, Yang Song, and Erik Meijering, 'Edge deep learning in computer vision and medical diagnostics: a comprehensive survey', *Artificial Intelligence Review*, vol. 58, article no. 93, Jan 2025, https://doi.org/10.1007/s10462-024-11033-5.
[112] Vishesh Narendra Pamadi, and Pushpa Singh, 'Edge AI vs Cloud AI: A Comparative Study of Performance Latency and Scalability', *International Journal of Research in Modern Engineering & Emerging Technology (IJRMEET)*, vol. 13, no. 3, pp. 13-35, Mar 2025, https://doi.org/10.63345/ijrmeet.org.v13.i3.2.
[113] Fernando Arias, Mayteé Zambrano Núñez, Ariel Guerra-Adames, Nathalia Tejedor-Flores, and Miguel Vargas-Lombardo, 'Sentiment Analysis of Public Social Media as a Tool for Health-Related Topics', *IEEE Access*, vol. 10, pp. 74850-74872, Jun 2022, https://doi.org/10.1109/ACCESS.2022.3187406.
[114] Chris Bell, Ayoolu Olukemi, Peter Broklyn, 'Social Media Sentiment Analysis for Brand Reputation Management', *Preprints*, article no. 2024080029, Jul 2024, https://doi.org/10.20944/preprints202408.0029.v1.
[115] Huang Huang, Adeleh Asemi Zavareh, and Mumtaz Begum Mustafa, 'Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions', *IEEE Access*, vol. 11, pp. 90367-90382, Aug 2023, https://doi.org/10.1109/ACCESS.2023.3307308
[116] Jiawen Li, Yuesheng Huang, Yayi Lu, Leijun Wang, Yongqi Ren, and Rongjun Chen, 'Sentiment Analysis Using E-Commerce Review Keyword-Generated Image with a Hybrid Machine Learning-Based Model', *Computers, Materials and Continua*, vol. 80, no. 1, pp. 1581-1599, Jul 2024, https://doi.org/10.32604/cmc.2024.052666.
[117] Shruti Thakur, and Sandeep Rajan, 'Sentiment Analysis on Customer Reviews of E-commerce Site', *M. Tech. Thesis*, Mar 2024.
[118] Shruti Gupta, and Raghav Sandhane, 'Use of sentiment analysis in social media campaign design and analysis', *Cardiometry*, Issue 22, May 2022; p. 351-363. Available from: http://www.cardiometry.net/issues/no22-may-2022/use_sentiment_analysis. https://doi.org/10.18137/cardiometry.2022.22.351363.
[119] Sulaxan Jadhav, Bahubali Shiragapur, Richa Purohit, and Nishant Jha, 'A Systematic Review on the Role of Sentiment Analysis in Healthcare', *Authorea*, Sep. 2024. Accepted for publication in *Applied AI Letters* on Jan 2025, https://doi.org/10.22541/au.172516418.81416769/v1.
[120] Omer S. Alkhnbashi, Rasheed Mohammad, and Mohammad Hammoudeh, 'Aspect-Based Sentiment Analysis of Patient Feedback Using Large Language Models', *Big Data and Cognitive Computing*, vol. 8, no. 12, article no. 167, Nov 2024, https://doi.org/10.3390/bdcc8120167
[121] Madan Mohan Tito Ayyalasomayajula, Akshay Agarwal, and Shahnawaz Khan, 'Reddit social media text analysis for depression prediction: using logistic regression with enhanced term frequency-inverse document frequency features', *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 5, pp. 5998-6005, Oct 2024, https://doi.org/10.11591/ijece.v14i5.pp5998-6005.
[122] Anushtha Vishwakarma, Mitali Chugh, 'COVID-19 vaccination perception and outcome: society sentiment analysis on twitter data in India', *Social Network Analysis and Mining*, vol. 13, no. 1, article no. 84, May 2023, https://doi.org/10.1007/s13278-023-01088-7.