Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Comprehensive Survey on Reliable Sentiment Analysis: Models, Datasets, and Future Directions

M. Amala Jeyaseeli[1], Dr. G. Heren Chellam[2]
[1]Research Scholar, Reg.No 22211172282007, Rani Anna Government College for Women
[2]Associate Professor, Rani Anna Government College for Women
[1,2]Affiliated to Manonmaniam Sundaranar University, Tirunelveli-627012
[1]amalafast@gmail.com, [2]hchellam@gmail.com

ARTICLE INFO

Received: 03 Nov 2024

Revised:20 Dec 2024

Accepted: 28 Dec 2024

ABSTRACT

Over the past decade, sentiment analysis has evolved from traditional rule-based and lexicon-driven approaches to advanced deep learning and transformer-based architectures capable of contextual and cross-lingual understanding. These advancements have expanded its applications across domains such as social media analytics, customer experience management, finance, and healthcare. However, challenges related to reliability, generalization, interpretability, and fairness remain significant. This study presents a comprehensive review of sentiment analysis from multiple perspectives, including datasets, modeling techniques, application domains, and evaluation frameworks. The paper examines traditional machine learning methods, deep learning architectures, and transformer-based models, highlighting their methodological foundations, advantages, and limitations. Particular attention is given to issues surrounding dataset quality, annotation strategies, ethical data collection, and bias in sentiment datasets. In addition, the study discusses critical challenges such as domain adaptation, multilingual sentiment analysis, sarcasm detection, and real-world deployment constraints. Emerging research directions and future opportunities are also outlined to support the development of more robust, interpretable, and ethically aligned sentiment analysis systems. Overall, this work provides a structured overview of current advancements and research gaps, serving as a reference framework for researchers and practitioners working on reliable sentiment analysis in natural language processing.

Keywords: Sentiment Analysis, Deep Learning, Transformer Models, Natural Language Processing (NLP), Opinion Mining.

1. Introduction

The rapid growth of user-generated content across social media, e-commerce platforms, news media, and blogs has created vast volumes of textual data reflecting public opinions and emotions. This development has increased the importance of Sentiment Analysis, also known as opinion mining, a key subfield of Natural Language Processing (NLP) that focuses on identifying, extracting, and classifying subjective information from textual data. Sentiment analysis enables organizations and researchers to monitor public opinion, evaluate customer satisfaction, and understand behavioural trends across various domains. Early sentiment analysis approaches relied primarily on lexicon-based methods, which used predefined dictionaries, linguistic rules, and sentiment word lists to determine polarity. Although effective in simple scenarios, these approaches were limited in capturing contextual nuances, linguistic variations, and complex semantic relationships. The introduction of machine learning techniques, including Naïve Bayes, Decision Trees, and Support Vector Machines, marked an important step toward automated sentiment classification. However, these models depended heavily on manually engineered features and often lacked generalization across domains, datasets, and languages.

Recent advances in deep learning have significantly improved sentiment analysis capabilities. Neural network architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) enable the modeling of sequential dependencies and hierarchical representations within text. Furthermore, the emergence of transformer-based architectures, including BERT, RoBERTa, GPT, and XLNet, has transformed the field by enabling contextualized language understanding at scale. These models have achieved state-of-the-art performance across various sentiment analysis tasks, including cross-lingual, aspect-based, and fine-grained sentiment classification. Despite these advancements, several challenges remain. Issues such as domain adaptation, dataset imbalance, cultural and contextual variations, sarcasm detection, model bias, and limited interpretability continue to affect the reliability and generalization of sentiment analysis systems. Moreover, many models rely heavily on benchmark datasets, which may restrict their applicability in real-world scenarios. As sentiment analysis becomes increasingly important in sectors such as business analytics, healthcare, governance, and social research, developing robust and trustworthy models has become essential.

In this context, the present study provides a structured overview of sentiment analysis research based on three major dimensions: models, datasets, and future research directions. The study traces the methodological evolution from lexicon-based approaches to modern transformer architectures and examines challenges related to cross-dataset, cross-domain, and multilingual reliability. The objective is to consolidate existing techniques, highlight research gaps, and identify potential directions for developing more reliable sentiment analysis systems.

The primary contributions of this study include:

1. **Unified Sentiment Analysis Taxonomy:** A comprehensive classification of sentiment analysis techniques, covering traditional methods, deep learning models, and transformer-based architectures.
2. **Datasets and Evaluation Metrics:** A review of widely used datasets, benchmarks, and evaluation metrics for sentiment analysis research.
3. **Reliability Analysis:** Examination of key factors influencing model reliability, including robustness, bias, interpretability, and ethical considerations.
4. **Future Research Opportunities:** Identification of emerging research directions, particularly in multimodal, multilingual, and explainable sentiment analysis.

The remainder of this paper is organized as follows. Section 2 presents the linguistic and computational foundations of sentiment analysis. Section 3 reviews major datasets and resources used for sentiment modeling. Section 4 discusses traditional machine learning approaches and feature engineering techniques. Section 5 examines recent advances in deep learning and transformer-based models. Section 6 highlights real-world applications across multiple domains. Section 7 focuses on evaluation methodologies and benchmarking frameworks. Section 8 outlines current research challenges, including bias, interpretability, and domain adaptation. Finally, Section 9 concludes the study by summarizing key findings and outlining future research directions.

2. Theoretical Background

Sentiment Analysis (SA) is a major research area within Natural Language Processing (NLP), data analytics, and machine learning that focuses on automatically identifying and classifying the emotional polarity of textual or spoken content. Early research treated sentiment analysis as a text classification task, associating words or phrases with positive, negative, or neutral sentiments [1], [5]. Initial approaches primarily relied on lexicon-based techniques, which used predefined sentiment dictionaries to determine opinion polarity. Although effective for simple scenarios, these methods often struggled to capture contextual nuances and complex emotional expressions in natural language [5].

With the rapid growth of user-generated content and social media data, rule-based methods proved insufficient, leading to the adoption of machine learning (ML) techniques. Supervised classifiers such as Naïve Bayes, Support Vector Machines (SVM), and Random Forests were widely used for sentiment classification using annotated corpora [2], [3]. These models improved classification performance

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

through engineered features such as n-grams, part-of-speech tags, and sentiment lexicons. However, ML-based approaches relied heavily on manual feature engineering and often showed limited generalization across domains [10].

The emergence of deep learning (DL) significantly advanced sentiment analysis research by enabling neural architectures to automatically learn hierarchical representations from raw text [3], [8]. Convolutional Neural Networks (CNNs) capture local contextual patterns, whereas Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models effectively model sequential dependencies in text. Empirical studies demonstrate that deep learning approaches outperform traditional ML methods in capturing semantic relationships and handling noisy or unstructured data [2], [8], [10]. In addition, hybrid architectures combining CNN and RNN components further enhance robustness across domains [6].

More recently, transformer-based architectures have further improved sentiment analysis performance. Models such as BERT [9], RoBERTa, XLNet, and GPT employ self-attention mechanisms to generate contextualized language representations. By leveraging large-scale pre-trained corpora, these models produce bidirectional contextual embeddings that enable deeper semantic understanding. As a result, transfer learning has become a dominant paradigm, allowing pre-trained language models to be fine-tuned for specific sentiment analysis tasks [9], [6]. Alongside these technological advances, increasing attention has been given to model interpretability and explain ability. Explainable Artificial Intelligence (XAI) methods aim to provide transparent explanations for model predictions, thereby enhancing trust, accountability, and ethical deployment of automated systems [4]. Sentiment analysis models are now widely applied across domains such as education, where student feedback is analyzed to evaluate learning experiences [7], and course evaluation systems that employ multitask learning frameworks [11].

Overall, sentiment analysis has evolved from rule-based lexicon approaches to advanced deep learning and transformer-based frameworks capable of capturing complex contextual and semantic relationships. These developments provide a strong conceptual and methodological foundation for modern sentiment analysis research and the development of more robust and interpretable systems.

3. Data Resources and Annotation Strategies

The performance of sentiment analysis models largely depends on the quality, diversity, and representativeness of the underlying datasets. Regardless of model sophistication, effectiveness is strongly influenced by the corpora used for training and evaluation. Consequently, research has shifted from small domain-specific datasets toward large-scale, multilingual, and context-aware corpora with richer annotation schemes, positioning dataset design as a central component of reliable sentiment analysis research [12].

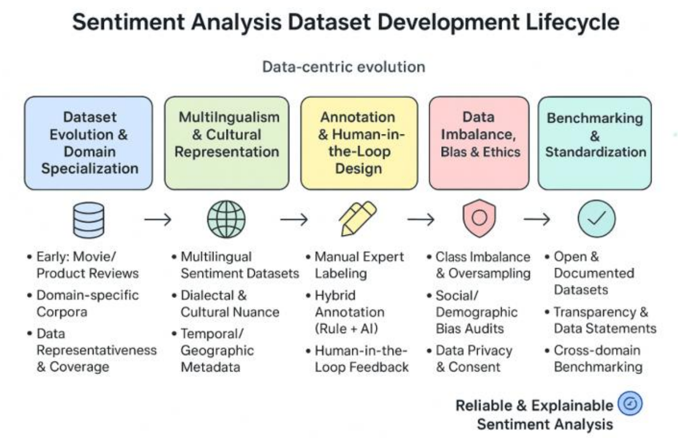

Figure 1. Sentiment Analysis Dataset Development Lifecycle

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Figure 1 shows the "Data life cycle of Sentiment Analysis" focusing on dataset evolution and multilingual representation, and then moving on to annotation, bias, and benchmarking. The focus is on making analysis ethical, diverse, and easy to explain, while data-gathering practices should be standard and explainable.

3.1 Dataset Evolution and Domain Specialization

Early sentiment analysis studies primarily relied on simple datasets such as movie reviews and rating-based corpora. While useful for initial experimentation, these datasets lacked linguistic diversity and domain depth. As the field evolved, researchers developed domain-specific corpora to better capture contextual and sentiment variations. For example, social media texts—characterized by short messages, emojis, and informal expressions—require different pre-processing and feature extraction strategies compared with formal reviews [14]. Similarly, specialized datasets in renewable energy discourse and healthcare narratives have shown improved capability in capturing socio-political and implicit sentiments [15], [22], [16].

However, the expansion of domain-specific datasets introduces domain adaptation challenges, as models trained on general corpora often exhibit reduced performance in specialized domains. Therefore, representative datasets incorporating diverse writing styles, topics, and sentiment intensities are essential for developing robust sentiment analysis systems [20].

3.2 Multilingualism and Cultural Representation

The globalization of digital communication has increased the demand for multilingual sentiment datasets. Research highlights the importance of datasets capable of capturing linguistic diversity, including code-switching, transliteration, and culturally dependent expressions [17], [18], [23]. Multilingual customer review datasets further support cross-lingual transfer learning, improving the performance of multilingual sentiment classification models [19].

Beyond linguistic diversity, cultural and temporal factors also influence sentiment interpretation. Emotional associations with words may vary across cultures and evolve over time. To address these dynamics, researchers increasingly incorporate auxiliary metadata—such as timestamps, user information, and geolocation—to analyze sentiment patterns across regions and time periods [14].

3.3 Annotation Techniques and Human-in-the-Loop Design

Reliable sentiment datasets require accurate and consistent annotation. Manual annotation remains the most dependable approach for generating gold-standard datasets, although it is labor-intensive [12]. To address scalability challenges, researchers increasingly employ semi-automated and hybrid annotation methods that combine heuristic techniques with expert supervision [15], [17].

Annotation quality is commonly evaluated using inter-annotator agreement metrics, such as Cohen's Kappa and Krippendorff's Alpha. Discrepancies typically arise from subjective interpretations, ambiguous expressions, or sarcasm [13]. Recently, Explainable Artificial Intelligence (XAI) and human-in-the-loop frameworks have been integrated into annotation workflows to support annotators, refine model predictions, and improve labeling consistency in large-scale datasets.

3.4 Data Imbalance, Bias, and Ethical Considerations

Dataset construction also faces challenges related to class imbalance, where positive and neutral sentiments frequently outnumber negative instances. Such imbalance can bias models toward dominant classes and reduce classification reliability [21]. Techniques such as oversampling, synthetic data generation, and data augmentation are commonly applied to mitigate this issue, although they must be carefully implemented to avoid introducing artificial patterns [24].

Bias and fairness concerns also remain critical. Underrepresentation of certain demographic groups may reinforce stereotypes and distort analytical outcomes. Consequently, recent studies emphasize the importance of fairness audits and transparent documentation of dataset characteristics [18], [14]. In addition, ethical considerations such as privacy protection, anonymization, and compliance with data protection regulations—including GDPR—are essential when collecting and processing social media data [15].

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

3.5 Benchmarking and Standardization

The growing complexity of sentiment analysis research highlights the importance of benchmark datasets and standardized evaluation frameworks. Public repositories integrating datasets from multiple domains—such as healthcare, finance, social media, and product reviews—enable comparative and reproducible research [12], [23].

Moreover, journals and conferences increasingly require detailed data statements describing data collection procedures, annotation methods, and potential biases. This emphasis on transparency promotes the development of reliable and ethically responsible sentiment analysis systems.

In summary, reliable sentiment analysis research depends on well-designed datasets, rigorous annotation strategies, and ethical data management. The increasing emphasis on domain-specific corpora, multilingual resources, human-in-the-loop annotation, and transparent benchmarking reflects a broader shift toward data-centric approaches in sentiment analysis, supporting the development of more robust and trustworthy models.

4. Sentiment Analysis Algorithms

Sentiment analysis algorithms aim to identify and classify emotions expressed in textual data. The process generally involves several stages, including text pre-processing, feature extraction, and classification. Over time, sentiment analysis techniques have evolved from traditional machine learning approaches to deep learning and transformer-based architectures.

4.1 Traditional Machine Learning Approaches

Before the emergence of deep learning, traditional machine learning (ML) methods were widely used for sentiment analysis. These supervised algorithms classify textual data into sentiment categories such as positive, negative, or neutral based on labeled datasets. Despite their relatively simple design, traditional ML methods established important foundations for feature engineering, reproducibility, and evaluation metrics, which remain relevant today.

Early ML-based sentiment analysis relied heavily on transforming text into numerical representations using techniques such as Bag-of-Words (BoW), TF–IDF weighting, n-grams, and part-of-speech features**.** Studies have shown that syntactic features such as unigrams and bigrams significantly improve sentiment classification performance, particularly in informal text such as social media data [28]. Furthermore, aspect-level sentiment analysis emerged as an important extension, enabling sentiment detection at the level of individual product attributes rather than entire documents, which is particularly valuable for e-commerce applications [27].

4.1.1 Probabilistic Methods: Naïve Bayes

The Naïve Bayes (NB) classifier is one of the earliest algorithms applied to sentiment analysis due to its probabilistic simplicity and computational efficiency. Studies demonstrate that NB performs effectively on large-scale datasets and noisy social media text, especially when parallelized for real-time analysis [25], [34]. However, NB often struggles with highly correlated features and long contextual dependencies, which can reduce accuracy in complex sentiment tasks [32]. Despite these limitations, its low computational cost and interpretability make NB a strong baseline for sentiment classification.

4.1.2 Linear Classifiers and Margin-Based Models

Support Vector Machines (SVMs) have proven particularly effective for high-dimensional textual data due to their ability to construct optimal decision boundaries in sparse feature spaces. Research shows that SVM-based models often outperform probabilistic classifiers such as NB, especially when kernel optimization techniques are applied [33]. Both linear and radial basis function (RBF) kernels have demonstrated strong performance depending on feature scale and dataset characteristics [29], [30]. In some studies, combining SVM with logistic regression and decision tree ensembles has further improved sentiment classification accuracy [26].

4.1.3 Ensemble and Hybrid Models

To address the limitations of individual classifiers, researchers have explored ensemble and hybrid frameworks that combine multiple algorithms. Ensemble approaches have demonstrated improved

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

recall and prediction accuracy compared with single-model systems, particularly in social media sentiment analysis tasks [37], [36]. Hybrid approaches, such as combining Naïve Bayes with SVM, have also shown improved performance in product sentiment evaluation [26]. These strategies laid the foundation for later multi-stage learning architectures used in modern sentiment analysis systems.

4.1.4 Domain and Language Adaptation

Traditional ML techniques have also been applied to domain-specific and multilingual sentiment analysis tasks. For instance, domain-dependent lexicons have been integrated with ML algorithms to analyze airline service reviews [31], while improved variants of Naïve Bayes have been developed for Arabic Twitter sentiment analysis [39]. Other studies demonstrate that careful feature selection and domain-specific data preparation significantly improve sentiment classification performance in specialized contexts [41].

Despite their contributions, traditional ML approaches face several limitations, including the need for manual feature engineering, limited ability to capture deep contextual relationships such as sarcasm and irony, and scalability challenges with extremely large datasets. Nevertheless, these models remain valuable in scenarios requiring interpretability, computational efficiency, and limited training data.

Overall, traditional machine learning methods established the fundamental principles of statistical modeling and feature-based sentiment analysis. Although deep learning models now dominate in terms of accuracy and scalability, classical ML techniques continue to play an important role in hybrid systems and serve as strong baselines for sentiment analysis research.

Table 1. Comparative View of Traditional ML Methods

Study / Year	Algorithm(s)	Dataset / Domain	Feature Techniques	Highlights / Key Findings
Jung et al. [25]	Enhanced Naïve Bayes	Twitter Streams	Bag-of-Words,	Achieved real-time classification with high throughput
Athindran [26]	Hybrid NB + SVM + Decision Trees	Brand Sentiment	TF-IDF, hybrid ensemble	Improved precision across competing brand analysis
Vanaja [27]	SVM (Aspect-level)	E-commerce Reviews	Aspect extraction, n-grams	Effective fine-grained sentiment detection
Iqbal et al. [28]	NB + Feature Combinations	Mixed Twitter Corpus	POS + N-gram fusion	Demonstrated synergy between syntactic and lexical features
Rathi et al. [29]	SVM, K-NN, NB	General Twitter Data	TF-IDF	Kernel selection crucial for optimal F1-scores
Rahat et al. [32]	NB vs SVM	Product Reviews	TF-IDF	NB better on small datasets; SVM scales better
Makhmudah [33]	SVM	Indonesian Tweets	Word frequency vectors	High accuracy on multilingual domain
Prabhakar et al. [37]	AdaBoost Ensemble	Airline Tweets	Combined base learners	Boosting improved recall and robustness
AlSalman [39]	NB, SVM	Arabic Tweets	Stemming, dialect normalization	Effective sentiment detection in morphologically rich language

4.2 Deep Learning and Transformer-Based Models

Over the past decade, sentiment analysis has significantly evolved with the emergence of deep learning and transformer-based architectures, shifting from manual feature engineering to representation learning. Unlike traditional machine learning methods that rely on static features, deep neural networks automatically learn hierarchical representations that capture contextual, semantic, and emotional

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

patterns in text [46], [47]. These advancements have substantially improved sentiment classification performance and expanded its applications across domains such as finance, healthcare, education, and social media analytics [42], [45].

4.2.1 Neural Feature Learning and End-to-End Modeling

Deep learning models enabled the transition from handcrafted features to end-to-end neural sentiment modeling. Convolutional Neural Networks (CNNs) effectively capture local semantic patterns and have demonstrated strong performance in social media sentiment detection, including sarcasm and multi-emotion expressions [44]. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks further improved sentiment modeling by capturing sequential dependencies and contextual relationships within sentences [49], [53]. These models have shown strong performance in tasks such as customer review analysis and healthcare feedback mining.

4.2.2 Bidirectional and Hybrid Architectures

To improve contextual understanding, bidirectional models such as BiLSTM and GRU were introduced to capture information from both past and future contexts. Hybrid architectures combining CNN, LSTM, and attention mechanisms have further enhanced sentiment detection accuracy and contextual representation [68], [66]. Such architectures improve the ability of neural models to capture subtle emotional patterns and semantic relationships within text.

4.2.3 Emergence of Transformer Models

The introduction of transformer architectures marked a major advancement in sentiment analysis. Transformers rely on self-attention mechanisms, enabling models to capture long-range dependencies and contextual relationships more efficiently than recurrent architectures [45]. The Bidirectional Encoder Representations from Transformers **(**BERT**)** model introduced contextual embeddings that significantly improved sentiment classification performance. Subsequent variants such as RoBERTa, ALBERT, DistilBERT, and GPT further enhanced efficiency and transfer learning capabilities. Recent research has also explored lightweight architectures such as FNet, which replaces attention with Fourier transforms while maintaining competitive performance in sentiment classification tasks [61]. In addition, multilingual transformer models have demonstrated strong cross-lingual sentiment classification capabilities when fine-tuned on domain-specific datasets [65].

4.2.4 Explainable and Multimodal Sentiment Analysis

As deep architectures become increasingly complex, model interpretability has become a critical research focus. Techniques such as attention visualization, gradient-based saliency maps, and layer-wise relevance propagation are widely used to interpret neural model predictions [49], [42]. Additionally, the integration of multimodal data—including text, audio, images, and video—has enabled more comprehensive sentiment analysis frameworks. For instance, multimodal sentiment models can combine linguistic and visual features to improve emotion detection in social media contexts [43].

4.2.5 Domain-Specific Deep Learning Applications

Deep learning and transformer models have enabled sentiment analysis applications across diverse domains. In finance, transformer-based models such as BERT and FinBERT have been applied to market sentiment prediction and investor behaviour analysis [45]. In healthcare, sentiment mining of patient feedback and clinical surveys supports service improvement and policy development [55], [54]. Similar approaches have been applied in public policy analysis, social media monitoring, and recommendation systems in e-commerce and tourism [51], [64], [65].

Despite these advances, several challenges remain, including data scarcity in low-resource languages, high computational costs, and potential bias in large models. Domain adaptation also remains difficult, as models trained on one domain often perform poorly when applied to another. Furthermore, the use of large generative models raises concerns regarding transparency, hallucination, and ethical data usage [56]. Ongoing research on efficient architectures, parameter-efficient tuning techniques such as LoRA**,** and model compression methods aims to improve scalability and sustainability in sentiment analysis systems.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

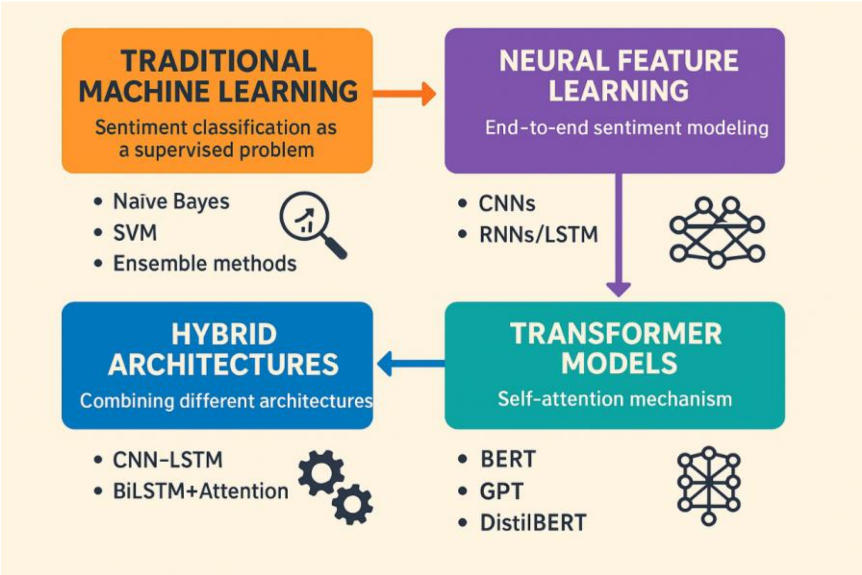

Figure 2. The Evolution of Sentiment Analysis Models

This figure illustrates a paradigm shift from traditional machine learning (Naïve Bayes, SVM) to neural and hybrid models, followed by transformer models (BERT, GPT). This shows a shift towards focusing on self-attentive, end-to-end sentiment models, providing a deeper context for understanding.

The evolution from shallow classifiers to deep neural and transformer models represents a major paradigm shift in sentiment analysis. Early architectures such as CNNs and LSTMs demonstrated the effectiveness of distributed representations, while transformer models introduced large-scale bidirectional contextual understanding. Modern sentiment analysis systems increasingly integrate explainability, multimodal data, and domain adaptation. Ongoing research on efficient and interpretable transformer architectures continues to advance sentiment analysis toward reliable and human-centered AI systems.

Table 2. Comparative summary of Deep and Transformer Models

Model / Architecture	**Core Mechanism**	**Representative Studies**	**Application Domain**	**Key Insights**
CNN	Convolutional filters for local feature extraction	Sailunaz & Alhajj (2019)	Twitter, product reviews	Strong for short text; limited context depth
LSTM / GRU	Sequential memory of word dependencies	Adak et al. (2022)	Food delivery, e-commerce	Captures temporal patterns; suffers from long sequences
BiLSTM	Bidirectional sequential encoding	Rana et al. (2023)	E-commerce recommendations	Improves contextual recall and sentiment gradient detection
Transformer (BERT, GPT)	Self-attention contextual embedding	Du et al. (2024); Al Maruf et al. (2024)	Finance, healthcare, general NLP	State-of-the-art semantic comprehension
FNet	Fourier transform in	Bhowmik et al. (2024)	Hospitality reviews	Lightweight; reduces computational cost

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

	place of attention			
Multimodal Transformer	Fusion of text and images	Zhang et al. (2021)	Social-media event analysis	Handles cross-modal emotion patterns
XAI-Integrated Transformer	Attention visualization and interpretability	Adak et al. (2022)	Customer feedback	Builds user trust and model transparency
Domain-Specific Transformer	Fine-tuned BERT/	Li et al. (2023);	Tourism, restaurants	High cross-lingual and contextual adaptability

5. Sentiment Analysis Datasets

Benchmark datasets play a crucial role in training and evaluating sentiment analysis models. While some studies rely on custom datasets, widely used public datasets provide standardized resources for model development and comparison. Common benchmarks include IMDb, Twitter US Airline Sentiment, Sentiment140, and SemEval-2017 Task 4, which represent different domains and linguistic characteristics.

Internet Movie Database (IMDb)

The IMDb dataset contains 50,000 movie reviews, evenly split into 25,000 training and 25,000 testing samples, with balanced positive and negative sentiments [98]. Due to the complex language and mixed narrative style, where reviews combine personal opinions with plot descriptions, the dataset is widely used as a benchmark for document-level sentiment classification.

Twitter US Airline Sentiment

The Twitter US Airline Sentiment dataset, collected by CrowdFlower, includes customer feedback on major US airlines with three sentiment classes—positive, negative, and neutral. The dataset contains 2,363 positive, 9,178 negative, and 3,099 neutral tweets, showing significant class imbalance. The informal language, abbreviations, and noise typical of social media texts make this dataset challenging for sentiment analysis models.

Sentiment140

The Sentiment140 dataset, developed at Stanford University, consists of approximately 1.6 million tweets labeled as positive or negative sentiment using emoticons as distant supervision [99]. Because tweets are short, informal, and context-dependent, the dataset presents realistic challenges for sentiment classification while providing a large-scale benchmark for social media sentiment analysis.

SemEval-2017 Task 4

The SemEval-2017 Task 4 dataset is a widely used benchmark for social media sentiment analysis [100]. It includes multilingual data, particularly English and Arabic, and supports multiple evaluation tasks such as message polarity classification, topic-based sentiment analysis, and tweet sentiment quantification, making it valuable for evaluating models under diverse experimental settings.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

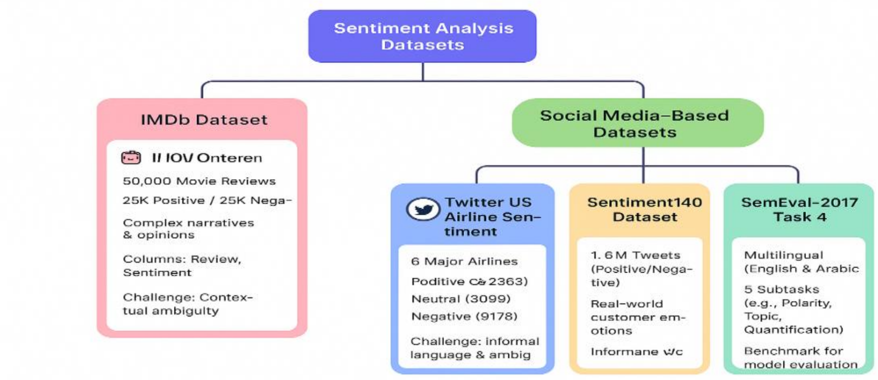

Figure 3. Sentiment Analysis Dataset

Figure 3 illustrates the division of sentiment analysis using the IMDB, Twitter US Airline Sentiment, Sentiment140, and SemEval-2017 Task 4 datasets. It emphasizes the context of language diversity and difficulty, such as amorphous and informal languages.

Table 3. Summary of Sentiment Analysis Datasets

Dataset	Classes	Strongly Positive	Positive	Neutral	Negative	Strongly Negative	Total
IMDb	2	-	25,000	-	25,000	-	50,000
Twitter US Airline Sentiment	3	-	2363	3099	9178	-	14,160
Sentiment140	2	-	800,000	-	800,000	-	1,600,000
SemEval-2017 4A	3	-	22,277	28,528	11,812	-	62,617
SemEval-2017 4B	2	-	17,414	-	7735	-	25,149
SemEval-2017 4C	5	1151	15,254	19,187	6943	476	43,011
SemEval-2017 4D	2	-	17,414	-	7735	-	25,149
SemEval-2017 4E	5	1151	15,254	19,187	6943	476	43,011

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 3 summarizes the key characteristics of these datasets, providing researchers with a comparative overview of commonly used benchmarks in sentiment analysis research.

6. Applications of Sentiment Analysis

Sentiment Analysis (SA) has evolved from an academic research topic into a widely adopted technology across multiple domains. By transforming unstructured textual opinions into structured insights, SA supports data-driven decision making and large-scale opinion monitoring [84], [85], [72].

- In business and marketing, SA is widely used for brand reputation monitoring, customer feedback analysis, and market trend prediction. Social media sentiment tracking enables organizations to evaluate audience reactions and optimize marketing strategies in real time [81]. Similarly, contextual sentiment models applied to online reviews help businesses improve services and customer satisfaction in retail and hospitality sectors [83], [76].

- In education and organizational analytics, SA helps evaluate learner engagement in online learning environments and analyze employee feedback within organizations. Emotion recognition techniques support improvements in course design and remote learning experiences (Chen et al., 2018), while sentiment mining of employee evaluations assists managerial decision-making and workplace monitoring [80], [74].

- SA is also applied in smart systems and social media analytics. Within the Social Internet of Things (SIoT), models such as bidirectional GRU networks enable sentiment-aware smart devices [72]. Social media sentiment analysis further supports public opinion monitoring, crisis detection, and misinformation analysis [87] although vulnerabilities such as adversarial attacks on deep models have been reported [79].

- In healthcare and public policy, sentiment mining of patient feedback, public forums, and government portals helps identify emotions such as trust, concern, and dissatisfaction. These insights support evidence-based policy decisions and service improvements [84].

Overall, sentiment analysis has evolved beyond simple polarity detection to become an important tool for understanding emotions, opinions, and behavioral patterns across diverse application domains, contributing significantly to human-centered AI and affective computing.

Table 4. Major application domains of sentiment analysis and their typical use cases.

Application Domain	Purpose of Sentiment Analysis	Typical Data Sources	Example Applications
Business & Marketing	Brand reputation monitoring and customer behavior analysis	Product reviews, social media posts	Social media campaign evaluation, customer satisfaction analysis
Education	Evaluation of learner engagement and course quality	Student feedback, online learning forums	Course improvement and adaptive learning systems
Human Resources	Workplace sentiment monitoring and employee feedback analysis	Employee surveys, internal communication	Organizational satisfaction and retention analysis
Internet of Things (IoT)	Emotion-aware smart systems and user interaction analysis	Smart device interactions, user commands	Sentiment-aware smart homes and intelligent assistants
Social Media Analytics	Public opinion monitoring and trend analysis	Tweets, online discussions	Crisis monitoring and misinformation detection
Healthcare & Policy	Public health perception and policy feedback analysis	Patient reviews, health forums	Healthcare service improvement and policy evaluation

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

7. Evaluation Frameworks and Benchmarks

Reliable evaluation is essential for assessing the performance of sentiment analysis systems. Early studies by Sebastiani (2001) and Pang et al. (2002) established benchmark-based evaluation using datasets such as IMDb and the Stanford Sentiment Treebank, enabling comparisons among classical classifiers including Naïve Bayes, SVM, and logistic regression. Later benchmarks expanded to domain-specific datasets such as Amazon reviews, Twitter, and SemEval, supporting aspect-level and cross-domain sentiment analysis. Modern evaluation frameworks use multiple metrics beyond accuracy, including precision, recall, F1-score, AUC–ROC, and confusion matrices, while also considering factors such as computational efficiency and latency for large models [90]. Deep learning models, particularly RNN-based architectures and transformer models such as BERT, have set new performance benchmarks through contextual representation learning [92], [93]. Recent research further extends evaluation to multilingual, multimodal, and cross-domain settings, emphasizing reproducibility, fairness, and interpretability [95]. Benchmarking platforms such as SentiBench and SentEval provide standardized evaluation pipelines across datasets and architectures. Overall, modern evaluation frameworks focus not only on accuracy but also on reliability, transparency, and ethical robustness of sentiment analysis systems.

8. Existing Challenges

Despite significant progress, sentiment analysis still faces several challenges that limit the development of robust and generalizable models. A major issue is data quality and domain bias, as many datasets are English-centric and domain-specific, often exhibiting class imbalance that affects model generalization. Even advanced transformer models such as BERT and GPT struggle with low-resource languages and specialized domains. Another challenge is the detection of sarcasm, irony, and evolving linguistic expressions, particularly in social media contexts where language usage changes rapidly. Additionally, model interpretability remains a critical concern, as deep learning and transformer models often function as black-box systems, raising accountability issues in sensitive domains such as healthcare, finance, and governance [95], [90]. Multimodal sentiment analysis involving text, audio, and video also introduces difficulties in feature alignment and information fusion. Furthermore, the increasing use of generative models raises concerns related to bias, privacy, and ethical accountability. Addressing these challenges requires the integration of explainable AI, cross-lingual learning, fairness-aware modeling, and standardized auditing frameworks to ensure reliable and transparent sentiment analysis systems.

9. Conclusion and Future Work

Sentiment analysis has evolved from lexicon-based and classical machine learning approaches to advanced deep learning and transformer-based architectures that capture contextual and semantic nuances in language. Traditional models such as Naïve Bayes and SVM laid the foundation for feature-based sentiment classification, while neural architectures—including CNNs, RNNs, and attention-based models—enabled deeper contextual understanding. More recently, transformer models such as BERT, RoBERTa, and GPT have significantly improved performance through contextual embeddings and transfer learning. Despite these advances, challenges such as linguistic ambiguity, sarcasm detection, data imbalance, cultural bias, and limited interpretability remain open research problems. Future sentiment analysis systems should focus on adaptive, transparent, and ethically responsible frameworks that work across multiple languages, domains, and modalities. Emerging directions include multimodal sentiment analysis, cross-lingual transfer learning, federated learning, and explainable AI (XAI) to enhance transparency and fairness. Additionally, improving computational efficiency through parameter-efficient tuning, pruning, and model compression will be essential for scalable deployment. Ultimately, the goal is to develop human-centered sentiment analysis systems that accurately interpret emotions while maintaining transparency, fairness, and social responsibility.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

References

[1] M. Ali, "NLTK Sentiment Analysis Tutorial for Beginners," *DataCamp* , 2023.

[2] L. Ashbaugh and Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers* , vol. 13, no. 12, p. 340, 2024.

[3] M. S. Basarslan and F. Kayaalp , "Sentiment Analysis on Social Media Reviews Datasets with Deep Learning Approach," *Sakarya Univ. J. Comput. Inf. Sci.* , vol. 4, no. 1, pp. 35–49, 2021.

[4] P. Biecek and W. Samek, "Position: Explain to Question Not to Justify," in *Proc. 41st Int. Conf. on Machine Learning* , 2024.

[5] V. Bonta, N. Kumaresh, and N. Janardhan, "A Comprehensive Study on Lexicon-Based Approaches for Sentiment Analysis," *Asian J. Comput. Sci. Technol.* , vol. 8, pp. 1–6, 2019.

[6] D. K. Dake and E. Gyimah, "Using Sentiment Analysis to Evaluate Qualitative Students' Responses," *Education and Information Technologies* , vol. 28, no. 4, pp. 4629–4647, 2023.

[7] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," *Electronics* , vol. 9, no. 3, 2020.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in *Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* , pp. 4171–4186, 2019.

[9] A. Ganie and S. Dadvandipour , "Traditional or Deep Learning for Sentiment Analysis: A Review," *Multidiszciplináris Tudományok* , vol. 12, no. 1, pp. 3–12, 2022.

[10] M. Birjali , M. Kasri, and A. Beni-Hssane , "A Comprehensive Survey on Sentiment Analysis: Approaches, Challenges and Trends," *Knowledge-Based Systems* , 2021.

[11] D. Kanojia and A. Joshi, "Applications and Challenges of Sentiment Analysis in Real-Life Scenarios," 2023.

[12] M. Rodríguez-Ibáñez, A. Casánez-Ventura, F. Castejón-Mateos , and P. M. Cuenca-Jiménez, "A Review on Sentiment Analysis from Social Media Platforms," *Expert Systems with Applications* , 2023.

[13] M. Kasri, A. El-Ansari, M. El Fissaoui, B. Cherkaoui, M. Birjali , and A. Beni-Hssane , "Public Sentiment Toward Renewable Energy in Morocco: Opinion Mining Using a Rule-Based Approach," *Social Network Analysis and Mining* , 2023.

[14] A. Rajput, "Natural Language Processing, Sentiment Analysis, and Clinical Analytics," *Innovations in Health Informatics: Smart Healthcare Primer* , 2019.

[15] M. Hankar , M. Birjali , A. El-Ansari, and A. Beni-Hssane , "COVID-19 Impact Sentiment Analysis on a Topic-Based Level," *Journal of ICT Standardization* , 2022.

[16] M. Hankar , M. Birjali , A. El-Ansari, and A. Beni-Hssane , "Arabic Topic Modeling -Based Sentiment Analysis on COVID-19 Feedback Comments," *Lecture Notes in Networks and Systems* , 2022.

[17] T. Quyyam and H. Ghous, "Sentiment Analysis of Amazon Customer Product Reviews: A Review," *International Journal of Scientific Research and Engineering Development* , 2021.

[18] N. M. Sharef, H. M. Zin, and S. Nadali, "Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data," *Journal of Computer Science* , 2016.

[19] P. P. Rokade , D. Aruna Kumari, "Business Intelligence Analytics Using Sentiment Analysis – A Survey," *International Journal of Electrical and Computer Engineering* , 2019.

[20] L. Abualigah , H. E. Alfar, M. Shehab, and A. M. A. Hussein, "Sentiment Analysis in Healthcare: A Brief Review," *Studies in Computational Intelligence* , 2020.

[21] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A Survey of Sentiment Analysis in Social Media," *Knowledge and Information Systems* , 2019.

[22] R. Bose, R. K. Dey, S. Roy, and D. Sarddar , "Sentiment Analysis on Online Product Reviews," *Advances in Intelligent Systems and Computing* , 2020.

[23] Jung, Y. G.; Kim, K. T.; Lee, B.; Youn, H. Y. "Enhanced Naive Bayes classifier for real-time sentiment analysis with SparkR." In *Proceedings of the 2016 IEEE International Conference on Information and Communication Technology Convergence (ICTC)* , Jeju Island, Republic of Korea, 19– October 21, 2016; pp. 141–146.

[24] Athindran, N. S.; Manikandaraj, S.; Kamaleshwar, R. "Comparative analysis of customer sentiments on competing brands using hybrid model approach." In *Proceedings of the 2018 IEEE 3rd International Conference on Inventive Computation Technologies (ICICT)* , Coimbatore, India, 15– November 16, 2018; pp. 348–353.

[25] Vanaja, S.; Belwal, M. "Aspect-level sentiment analysis on e-commerce data." In *Proceedings of the 2018 IEEE International Conference on Inventive Research in Computing Applications (ICIRCA)* , Coimbatore, India, 11– July 12, 2018; pp. 1275–1279.

[26] Iqbal, N.; Chowdhury, A. M.; Ahsan, T. "Enhancing the performance of sentiment analysis by using different feature combinations." In *Proceedings of the 2018 IEEE International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)* , Rajshahi, Bangladesh, 8– February 9, 2018; pp. 1–4.

[27] Rathi, M.; Malik, A.; Varshney, D.; Sharma, R.; Mendiratta, S. "Sentiment analysis of tweets using machine learning approach." In *Proceedings of the 2018 IEEE Eleventh International Conference on Contemporary Computing (IC3)* , Noida, India, 2– August 4, 2018; pp 1–3.

[28] Tariyal, A.; Goyal, S.; Tantububay, N. "Sentiment Analysis of Tweets Using Various Machine Learning Techniques." In *Proceedings of the 2018 IEEE International Conference on Advanced Computation and Telecommunication (ICACAT)* , Bhopal, India, 28– December 29, 2018; pp. 1–5.

[29] Hemakala , T.; Santhoshkumar, S. "Advanced classification method of Twitter data using sentiment analysis for airline service." *International Journal of Computer Science & Engineering* , 2018, vol. 6, pp. 331–335.

[30] Rahat, A. M.; Kahir, A.; Masum, A. K. M. "Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset." In *Proceedings of the 2019 IEEE 8th International Conference System Modeling and Advancement in Research Trends (SMART)* , Moradabad, India, 22– November 23, 2019; pp. 266–270.

[31] Makhmudah, U.; Bukhori, S.; Putra, J. A.; Yudha, B. A. B. "Sentiment Analysis of Indonesian Homosexual Tweets Using Support Vector Machine Method." In *Proceedings of the 2019 IEEE International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)* , Jember, Indonesia, 16– October 17, 2019; pp. 183–186.

[32] Wongkar, M.; Angdresey, A. "Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter." In *Proceedings of the 2019 IEEE Fourth International Conference on Informatics and Computing (ICIC)* , Semarang, Indonesia, 16– October 17, 2019; pp.1–5.

[33] Madhuri, D. K. "A machine learning based framework for sentiment classification: Indian railways case study." *International Journal of Innovative Technology & Exploring Engineering (IJITEE)* 2019, vol. 8, pp. 441–445.

[34] Gupta, A.; Singh, A.; Pandita, I.; Parashar, H. "Sentiment analysis of Twitter posts using machine learning algorithms." In *Proceedings of the 2019 IEEE 6th International Conference on Computing for Sustainable Global Development (INDIACom)* , New Delhi, India, 13– March 15, 2019; pp. 980–983.

[35] Prabhakar, E.; Santhosh, M.; Krishnan, A. H.; Kumar, T.; Sudhakar, R. "Sentiment analysis of US Airline Twitter data using new AdaBoost approach." *International Journal of Engineering Research & Technology (IJERT)* , 2019, vol. 7, pp. 1–6.

[36] Hourrane, O.; Idrissi, N. "Sentiment Classification on Movie Reviews and Twitter: An Experimental Study of Supervised Learning Models." In *Proceedings of the 2019 IEEE 1st International Conference on Smart Systems and Data Science (ICSSD)* , Rabat, Morocco, 3– October 4, 2019; pp. 1–6.

[37] AlSalman, H. "An Improved Approach for Sentiment Analysis of Arabic Tweets in Twitter Social Media." In *Proceedings of the 2020 IEEE 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, 19– March 21, 2020; pp 1–4.

[38] Saad, A. I. "Opinion Mining on US Airline Twitter Data Using Machine Learning Techniques." In *Proceedings of the 2020 IEEE 16th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt, 29– December 30, 2020; pp. 59–63.

[39] Alzyout, M.; Bashabsheh, E. A.; Najadat, H.; Alaiad, A. "Sentiment Analysis of Arabic Tweets about Violence Against Women Using Machine Learning." In *Proceedings of the 2021 IEEE 12th International Conference on Information and Communication Systems (ICICS)*, Valencia, Spain, 24– May 26, 2021.

[40] A. Al Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha, and Z. Aung, "Challenges and Opportunities of Text-Based Emotion Detection: A Survey," IEEE Access, vol. 12, pp. 18416–18450, 2024.

[41] H. Zhang, S. Qian, Q. Fang, and C. Xu, "Multimodal Disentangled Domain Adaptation for Social Media Event Rumor Detection," IEEE Trans. Multimedia, vol. 23, pp. 4441–4454, 2021.

[42] K. Sailunaz and R. Alhajj, "Emotion and Sentiment Analysis from Twitter Text," J. Comput. Sci., 2019.

[43] K. Du, F. Xing, R. Mao, and E. Cambria, "Financial Sentiment Analysis: Techniques and Applications," ACM Computing Surveys, vol. 56, no. 9, pp. 1–42, 2024.

[44] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and Sentiment Analysis: An Overview of the Current State of the Area and Envisaged Developments," Decision Support Systems, vol. 53, no. 4, pp. 675–679, 2012.

[45] A. Khan et al., "Sentiment Classification of User Reviews Using Supervised Learning Techniques with Comparative Opinion Mining Perspective," Advances in Intelligent Systems and Computing, vol. 944, pp. 23–29, 2020.

[46] T. A. Al-Qablan, M. H. Mohd Noor, M. A. Al-Betar, and A. T. Khader, A Survey on Sentiment Analysis and Its Applications. London: Springer; 2023.

[47] A. Adak, B. Pradhan, and N. Shukla, "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review," Foods, vol. 11, no. 10, p. 1500, 2022.

[48] P. Shah et al., "A Comprehensive Review on Sentiment Analysis of Social/Web Media Big Data for Stock Market Prediction," Int. J. Syst. Assur. Eng. Manag., vol. 15, no. 6, pp. 2011–2018, 2024.

[49] J. Lappeman, A. Goder, K. Naicker, H. Faruki, and P. Gordon, "Using Sentiment Analysis to Understand Public Policy Nicknames: Obamacare and the Affordable Care Act," J. Nonprofit Public Sect. Mark., vol. 36, no. 3, pp. 347–363, 2023.

[50] R. F. P. Aramburo, M. Â. L. Moreira, L. P. L. Fávero, I. P. De Araújo Costa, and M. Dos Santos, "Data Science in Social Politics with Particular Emphasis on Sentiment Analysis," Procedia Comput. Sci., vol. 214, pp. 420–427, 2022.

[51] R. M. M. Magtangob and T. D. Palaoag, "Assessment of the Healthcare Administration of Senior Citizens from Survey Data Using Sentiment Analysis," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 2, pp. 389–394, 2023.

[52] D. Georgiou, A. MacFarlane, and T. Russell-Rose, "Extracting Sentiment from Healthcare Survey Data: An Evaluation of Sentiment Analysis Tools," Proc. 2015 Sci. Inf. Conf. (SAI), 2015, pp. 352–361.

[53] M. Srisankar, "A Survey on Sentiment Analysis Techniques in the Medical Domain," Medicon Agric. Environ. Sci., vol. 6, no. 2, pp. 4–9, 2024.

[54] F. Sudirjo, K. Diantoro, J. A. Al Gasawneh, H. K. Azzaakiyyah, and A. M. Almaududi Ausat, "Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses," J. Teknol. Dan Sist. Inf. Bisnis, vol. 5, no. 3, pp. 283–288, 2023.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[55] M. Axhiu, F. Veljanoska, B. Ciglovska, and M. Husejni, "The Usage of Sentiment Analysis for Hearing the Voice of the Customer and Improving Businesses," J. Educ. Soc. Res., 2014.

[56] K. Anbazhagan, P. Singhal, M. Gupta, and K. Saxena, "Sentiment Analysis of Online Customer Feedbacks Using NLP and Supervised Learning Algorithm," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 3s, pp. 391–397, 2024.

[57] P. W. Cahyo, U. S. Aesyi, and B. D. Santosa, "Topic Sentiment Using Logistic Regression and Latent Dirichlet Allocation as a Customer Satisfaction Analysis Model," J. Infotel, vol. 16, no. 1, pp. 1–16, 2024.

[58] V. Jain, A., & Mitra, "Development and Application of Machine Learning Algorithms for Sentiment Analysis in Digital Manufacturing: A Pathway for Enhanced Customer Feedback," Emerging Technologies Digital Manufacturing Smart Factories, 2023.

[59] S. Bhowmik, R. Sadik, W. Akanda, and J. R. Pavel, "Sentiment Analysis with Hotel Customer Reviews Using FNet," Bull. Electr. Eng. Informatics, vol. 13, no. 2, pp. 1298–1306, 2024.

[60] A. Mawadati, W. Ustyannie, A. H. Wibowo, and R. A. Simanjuntak, "Analysis of Yogyakarta Coffee Shop Visitor Reviews to Increase Customer Satisfaction Using Sentiment Analysis," KnE Social Sciences, 2024, pp. 30–39.

[61] L. Bharadwaj, "Sentiment Analysis in Online Product Reviews: Mining Customer Opinions for Sentiment Classification," Int. J. Multidiscip. Res., 2023.

[62] H. Li, B. X. B. Yu, G. Li, and H. Gao, "Restaurant Survival Prediction Using Customer-Generated Content: An Aspect-Based Sentiment Analysis of Online Reviews," Tourism Management, vol. 96, 2023.

[63] S. Ounacer, D. Mhamdi, S. Ardchir, A. Daif, and M. Azzouazi, "Customer Sentiment Analysis in Hotel Reviews Through Natural Language Processing Techniques," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 1, pp. 569–579, 2023.

[64] S. Nazirkar and S. Kulkarni, "Sentiment Analysis and Customer Satisfaction Factors Based on LSTM and Topic Modeling," Indian J. Sci. Technol., vol. 16, no. 28, pp. 2126–2132, 2023.

[65] A. Zakaria and M. Siallagan, "Predicting Customer Satisfaction Through Sentiment Analysis on Online Review," Int. J. Curr. Sci. Res. Rev., vol. 6, no. 1, pp. 515–522, 2023.

[66] M. R. R. Rana, A. Nawaz, T. Ali, A. M. El-Sherbeeny, and W. Ali, "A BiLSTM-CF and BiGRU-Based Deep Sentiment Analysis Model to Explore Customer Reviews for Effective Recommendations," Eng. Technol. Appl. Sci. Res., vol. 13, no. 5, pp. 11739–11746, 2023.

[67] B. Andrian, T. Simanungkalit, I. Budi, and A. F. Wicaksono, "Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 3, pp. 466–473, 2022.

[68] A. S. Oliveira, A. I. Renda, M. B. Correia, and N. Antonio, "Hotel Customer Segmentation and Sentiment Analysis Through Online Reviews: An Analysis of Selected European Markets," Tourism Management Studies 18, no. 1, pp. 29–40, 2022.

[69] M. Ma'ruf, A. P. Kuncoro, P. Subarkah, and F. Nida, "Sentiment Analysis of Customer Satisfaction Levels on Smartphone Products Using Ensemble Learning," Ilk J. Ilm., vol. 14, no. 3, pp. 339–347, 2022.

[70] W. Ali, Y. Yang, X. Qiu, Y. Ke, and Y. Wang, "Aspect-Level Sentiment Analysis Based on Bidirectional-GRU in SIoT," IEEE Access, vol. 9, pp. 69938–69950, 2021.

[71] B. Amina and T. Azim, "SCANCPECLENS: A Framework for Automatic Lexicon Generation and Sentiment Analysis of Micro-Blogging Data on China-Pakistan Economic Corridor," IEEE Access, vol. 7, pp. 133876–133887, 2019.

[72] I. Awajan, M. Mohamad, and A. Al-Quran, "Sentiment Analysis Technique and Neutrosophic Set Theory for Mining and Ranking Big Data from Online Reviews," IEEE Access, vol. 9, pp. 47338–47353, 2021.

[73] Q. Bai, Q. V. Hu, L. Ge, and L. He, "Stories That Big Danmaku Data Can Tell as a New Media," IEEE Access, vol. 7, pp. 53509–53519, 2019.

[74] X. Cao, J. Yu, and Y. Zhuang, "Injecting User Identity Into Pretrained Language Models for Document-Level Sentiment Classification," IEEE Access, vol. 10, pp. 30157–30167, 2022.

[75] M.-H. Chen, W.-F. Chen, and L.-W. Ku, "Application of Sentiment Analysis to Language Learning," IEEE Access vol. 6, pp. 24433–24442, 2018.

[76] R. Cobos, F. Jurado, and A. Blázquez-Herranz, "A Content Analysis System That Supports Sentiment Analysis for Subjectivity and Polarity Detection in Online Courses," IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, vol. 14, no. 4, pp. 177–187, 2019.

[77] J. Dai, C. Chen, and Y. Li, "A Backdoor Attack Against LSTM-Based Text Classification Systems," IEEE Access, vol. 7, pp. 138872–138878, 2019.

[78] A. S. de Oliveira Góes and R. C. L. de Oliveira, "A Process for Human Resource Performance Evaluation Using Computational Intelligence: An Approach Using a Combination of Rule-Based Classifiers and Supervised Learning Algorithms," IEEE Access, vol. 8, pp. 39403–39419, 2020.

[79] S. Gupta and R. Sandhane, "Use of Sentiment Analysis in Social Media Campaign Design and Analysis," Cardiometry, no. 22, pp. 351–363, 2022.

[80] Z. Hameed and B. Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," IEEE Access, vol. 8, pp. 73992–74001, 2020.

[81] L. Huang, Z. Dou, Y. Hu, and R. Huang, "Textual Analysis for Online Reviews: A Polymerization Topic Sentiment Model," IEEE Access, vol. 7, pp. 91940–91945, 2019.

[82] D. M. E.-D. M. Hussein, "A Survey on Sentiment Analysis Challenges," J. King Saud Univ. – Eng. Sci., vol. 30, no. 4, pp. 330–338, 2018.

[83] F. Iqbal, J. M. Hashmi, B. C. M. Fung, R. Batool, A. M. Khattak, S. Aleem and P. C. K. Hung, "A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction," IEEE Access, vol. 7, pp. 14637–14652, 2019.

[84] M. Jiang, J. Wu, X. Shi, and M. Zhang, "Transformer-Based Memory Network for Sentiment Analysis of Web Comments," IEEE Access, vol. 7, pp. 179942–179953, 2019.

[85] R. Kamal, M. A. Shah, C. Maple, M. Masood, A. Wahid, and A. Mehmood, "Emotion Classification and Crowd Source Sensing: A Lexicon-Based Approach," IEEE Access, vol. 7, pp. 27124–27134, 2019.

[86] E. Kochmar, Getting Started with Natural Language Processing. Manning Publications, 2022.

[87] L. Li, Y. Wu, Y. Zhang, and T. Zhao, "Time + User Dual Attention Based Sentiment Prediction for Multiple Social Network Texts With Time Series," IEEE Access, vol. 7, pp. 17644–17653, 2019.

[88] R. Aggarwal, "How to Measure the Efficacy of Your Sentiment Analysis Model," SearchUnify Technical Blog, 2024, https://www.searchunify.com/blog/how-to-me [Online]. Available: https://www.searchunify.com/sudo-technical-blogs/how-to-measure-the-efficacy-of-your-sentimentanalysis-model/

[89] LinkedIn Editorial, "How Can You Evaluate Sentiment Analysis Models? LinkedIn Advisory Article, 2024. [Online]. Available: https://www.linkedin.com/advice/1/how-can-you-evaluate-sentiment-analysis-model-ygfec

[90] D. Tang, B. Qin, and T. Liu, "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification," Proc. Association for Computational Linguistics, 2015. [Online]. Available: http://ir.hit.edu.cn/

[91] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," Google AI Language Research, 2019. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[92] N. Karim and N. Habbat, "Customer Sentiment Analysis for Arabic Social Media Using a Novel Ensemble Machine Learning Approach," Int. J. Electr. Comput. Eng., vol. 13, no. 4, pp. 4504–4515, Aug. 2023.

[93] N. S. Kyritsis and M. P. K. Kyritsis, "A Comparative Performance Evaluation of Algorithms for the Analysis and Recognition of Emotional Content," Artificial Intelligence, IntechOpen, Jan. 2024.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[94] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," in Proc. Empirical Methods in Natural Language Processing, 2002. [Online]. Available: http://reviews.imdb.com/Reviews/

[95] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, 2001. [Online]. Available: http://liinwww.ira.uka.de/bibliography/Ai/automated.text.categorization.html

[96] Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets and future research. *Applied Sciences* , *13* (7), 4550.

[97] Wagh, R., & Punde, P. (2018, March). Survey on sentiment analysis using the Twitter dataset. In *2018 , the Second International Conference on Electronics , Communication and Aerospace Technology (ICECA)* (pp. 208-211). IEEE.