Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Sentiment Analysis Based on Machine Learning Techniques: A Comprehensive Review

Ari Ibrahim Hamid[1], Adnan Mohsin Abdulazeez [2]
ari.hamid@dpu.edu.krd[1], adnan.mohsin@dpu.edu.krd[2]
[1]IT department, Technical College of Duhok, Duhok Polytechnic University, Duhok, Iraq
[2]Technical College of Engineering, Duhok Polytechnic University, Duhok, Iraq

Article Information

Submitted: 20 May 2024
Reviewed: 23 May 2024
Accepted : 15 Jun 2024

Keywords

Sentiment Analysis, Machine Learning, Social Media Analysis

Abstract

In the landscape of digital communication, sentiment analysis stands out as a pivotal technology for deciphering the vast troves of unstructured text generated online. When integrated with machine learning, sentiment analysis transforms into a powerful tool capable of distilling insights from complex human emotions and opinions expressed across social media, reviews, and forums. This review paper embarks on a thorough exploration of the integration of machine learning techniques with sentiment analysis, shedding light on the latest advancements, challenges, and applications spanning various sectors including public health, finance, and consumer behavior. It meticulously examines the role of machine learning in elevating sentiment analysis through improved accuracy, adaptability, and depth of analysis. Furthermore, the paper discusses the implications of these technologies in understanding consumer sentiment, tracking public health trends, and forecasting market movements. By synthesizing findings from seminal studies and cutting-edge research, this review not only charts the current landscape but also forecasts the trajectory of sentiment analysis. It underscores the necessity for ongoing innovation in machine learning models to keep pace with the evolving digital discourse. The insights presented herein aim to guide future research endeavors, highlight the transformative impact of machine learning on sentiment analysis, and outline the potential for new applications that could benefit society at large.

A. Introduction

In the contemporary digital landscape, sentiment analysis emerges as a pivotal tool for interpreting the vast quantities of unstructured data generated across various online platforms. This domain, particularly when enhanced by machine learning techniques, offers profound insights into public perceptions, opinions, and emotions towards products, services, policies, and social phenomena. Our review paper, titled "Sentiment Analysis based on Machine Learning Techniques: A Comprehensive Review," aims to delineate the significant strides and methodological innovations in this field, as highlighted by a range of seminal works and recent studies. The advent of machine learning in sentiment analysis has transformed the ability to analyze large-scale user reviews, enabling more nuanced and accurate interpretations of sentiment. Al-Ghuribi et al. highlighted the efficacy of unsupervised semantic approaches in aspect-based sentiment analysis, showcasing the potential for analyzing extensive user reviews with heightened precision [1]. Similarly, the emergence of the COVID-19 pandemic catalyzed the application of sentiment analysis in monitoring public sentiment. Studies by Braig et al. and Costola et al. illustrated how machine learning techniques were pivotal in analyzing sentiments related to COVID-19 on platforms like Twitter and correlating them with stock market reactions [2], [3]. Moreover, the utility of sentiment analysis spans various domains, from detecting fake news to analyzing e-commerce product experiences and monitoring mental health through social media. For instance, Dev et al. developed a hybrid machine learning model that unmasked fake news, showcasing the cross-disciplinary applicability of sentiment analysis [4]. Furthermore, He et al. explored e-commerce product experiences through a fusion sentiment analysis method, emphasizing the importance of sentiment analysis in enhancing consumer experience [5]. Technological advancements have also facilitated the evolution of sentiment analysis techniques. Dai et al. employed unsupervised learning for sentiment analysis by leveraging multi-source knowledge transfer, illustrating the shift towards more sophisticated and nuanced analysis methods that can harness knowledge from diverse sources [6]. This evolution points to a broader trend of integrating complex machine learning models and algorithms to improve the accuracy and depth of sentiment analysis. The application of sentiment analysis in financial markets and health informatics further exemplifies its wide-reaching impact. Mehta et al. demonstrated how machine learning and sentiment analysis could be used for stock price prediction, highlighting its significance in financial decision-making [7]. Additionally, Kazmaier and van Vuuren underscored the power of ensemble learning in sentiment analysis, providing insights into the methodological diversity within the field [8]. The progression towards more advanced machine learning models is evident in the work of Mahalakshmi et al., who utilized a conditional generative adversarial network for Twitter sentiment analysis, marking a significant methodological advancement [9]. Similarly, the synergy between stock prices and investor sentiment in social media, as explored by Liu et al., underscores the practical implications and potential financial applications of sentiment analysis [10]. Emerging technologies and methodologies in sentiment analysis have facilitated the expansion of its applications. The exploration of sentiment toward emerging infectious diseases on social media by

Lee et al. further demonstrates the application of dictionary-based sentiment analysis in public health and safety [11]. The versatility of sentiment analysis is further exemplified in the domain of financial news, where Yadav et al. employed an unsupervised approach to sentiment analysis, highlighting the adaptability of machine learning techniques to various content types and contexts [12]. The integration of machine learning with sentiment analysis has not only enhanced the accuracy and efficiency of sentiment evaluations but also broadened the spectrum of its applicability, encompassing areas such as stock market forecasting by Ueda et al. and sentiment dynamics in physician rating websites by Shah et al., further emphasizing the multidisciplinary potential of sentiment analysis [13], [14]. The integration of machine learning with sentiment analysis has also led to novel applications that address specific societal and business needs. For example, Hinduja et al. harnessed machine learning for mental health monitoring using Twitter data, showcasing the potential of sentiment analysis in proactive health and wellbeing monitoring [15]. In addition to technological innovation, the research community has also explored the impact of sentiment analysis on various aspects of daily life and industry-specific applications. For instance, Li et al. used social media big data for tourist demand forecasting, employing a new machine learning analytical approach that underscores the importance of sentiment analysis in predicting consumer behavior trends [16]. The continuous evolution of sentiment analysis, driven by advancements in machine learning techniques, has opened new avenues for research and application. through a fusion sentiment analysis method and the analysis of public opinion on food safety in Greater China by Zhang et al. are indicative of the expanding scope of sentiment analysis applications, from consumer behavior to public health and safety [17]. As we advance into the nuances of sentiment analysis, it's crucial to recognize the diversity of its applications and the various machine learning techniques that enhance its accuracy and applicability. The development of Turkish sentiment analysis models by Demircan et al. using machine learning and e-commerce data highlights the linguistic and cultural adaptations required for effective sentiment analysis, showcasing the importance of localized models in global applications [18]. The intersection of sentiment analysis with health informatics presents another fascinating application area. Rybinski's work on predicting news longevity from textual and context features through machine learning models exemplifies how sentiment analysis can be extended beyond traditional domains, contributing to our understanding of information dissemination and its impact [19]. As the field of sentiment analysis continues to evolve, driven by advances in machine learning and computational linguistics, its applications extend into new and unexpected domains. The study of aspect-based sentiment analysis optimization by Marutho et al. using advanced techniques such as sentence embedding transformers and sparse attention mechanisms indicates the potential for further methodological advancements that could enhance the accuracy and applicability of sentiment analysis across various sectors [20].

The review paper is organized into several key sections: Section 2 delves into the relationship between machine learning and sentiment analysis, detailing how different machine learning techniques contribute to advancing sentiment analysis applications. Section 3 provides a comprehensive review of the related literature, with significant studies summarized in Table 1 to allow for easy comparison. The paper concludes with Section 4, which includes a discussion, the study's conclusions, and a detailed list of references.

B. Research Method

1. Introduction to Sentiment Analysis

Sentiment Analysis, also known as opinion mining, is a computational study of people's opinions, sentiments, attitudes, and emotions expressed in text. It is pivotal for interpreting the vast quantities of unstructured data generated across various online platforms. This domain, when enhanced by machine learning techniques, offers profound insights into public perceptions, opinions, and emotions towards products, services, policies, and social phenomena, making it a critical tool for businesses and researchers alike [21].

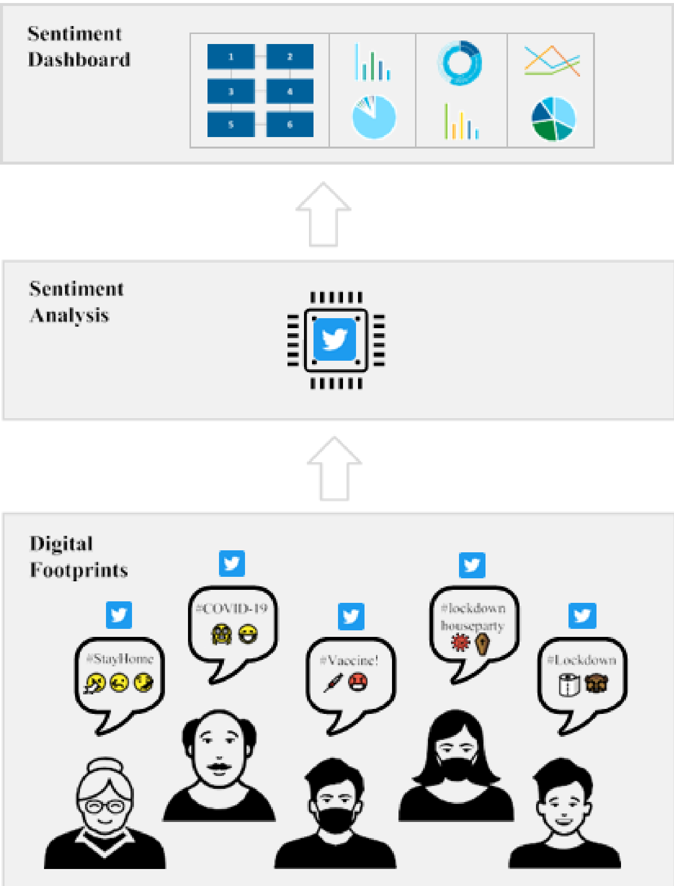

Figure 1. Sentiment analysis of COVID-19-related tweets can provide relevant insights for effectively managing the pandemic and its impacts [22].

2. Machine Learning in Sentiment Analysis

The integration of machine learning in sentiment analysis has transformed the field by enabling the automated analysis of large-scale data. This evolution from traditional algorithms to sophisticated machine learning techniques has significantly improved the accuracy and efficiency of sentiment analysis. These advancements have allowed for more nuanced interpretations of sentiment, making it possible to analyze and understand complex patterns of human emotion and opinion across diverse datasets [23].

3. Categories of Machine Learning Algorithms in Sentiment Analysis

Machine learning algorithms used in sentiment analysis can be categorized into supervised learning, unsupervised learning, and hybrid approaches. Each category has its unique advantages and applications, making them suitable for different aspects of sentiment analysis.

3.1 Supervised Learning Algorithms

Supervised learning algorithms in sentiment analysis involve the use of labeled datasets to train models that can classify or predict sentiments. These algorithms learn from the input-output pairs during the training phase, allowing them to make predictions or classifications on unseen data. Popular supervised learning algorithms include Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Random forest, Decision tree, and Naive Bayes, which have shown significant effectiveness in various sentiment analysis tasks [24].

3.1.1 Support Vector Machines (SVM)

Support Vector Machines (SVM) are a set of supervised learning methods used for classification, regression, and outliers detection. The core principle of SVM is to find the hyperplane in an N-dimensional space that distinctly classifies the data points. It is particularly useful in high-dimensional spaces and for cases where the number of dimensions exceeds the number of samples. This makes SVM suitable for applications like sentiment analysis and text classification, as demonstrated in studies analyzing news sentiment, financial market forecasts, and social media sentiment classification [25].

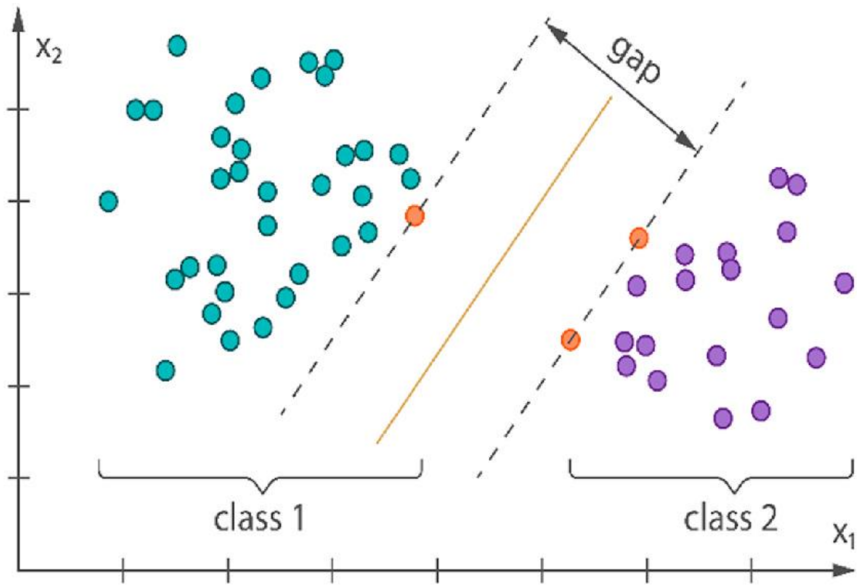

Figure 2. Support Vector Machines (SVM) [26].

3.1.2 K-Nearest Neighbor (KNN)

K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm where the class of a sample is determined by the majority class among its k nearest neighbors. KNN is non-parametric and versatile, being used for both classification and regression tasks. Its simplicity and effectiveness for tasks that involve finding similarities between instances make it applicable for sentiment analysis, such as classifying sarcastic sentiments in the aviation sector or early-stage pregnancy recognition on microblogs [27].

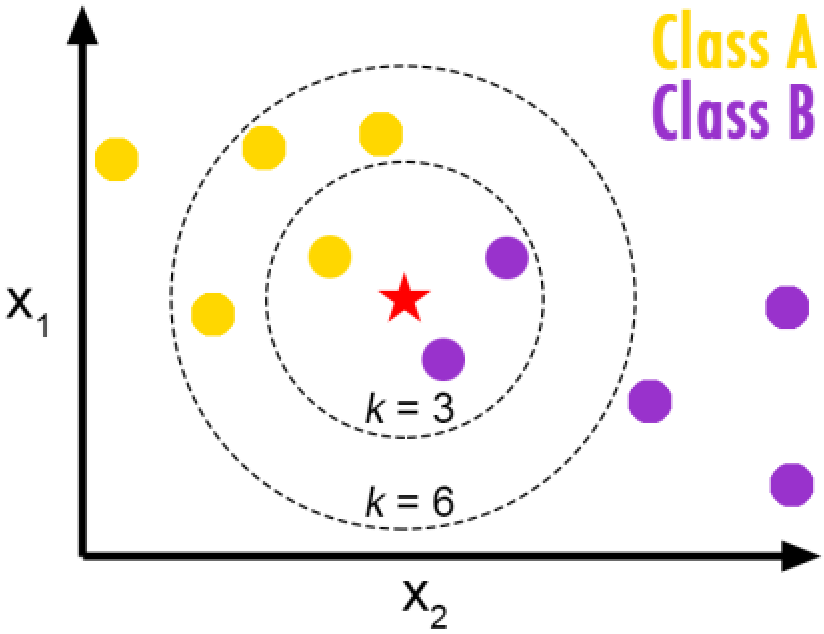

Figure 3. K- Nearest Neighbors (KNN) [28].

3.1.3 Random forest:

Decision trees are a non-linear predictive modeling tool used for classification and regression tasks. They mimic human decision-making processes by splitting data into branches at decision nodes, based on some decision criteria. Due to their simplicity and interpretability, decision trees are widely used in areas such as sentiment analysis, where they can help in analyzing text data for sentiment and intent, including in the contexts of financial stock market forecasts and analyzing the impact of press media on public sentiment [29].

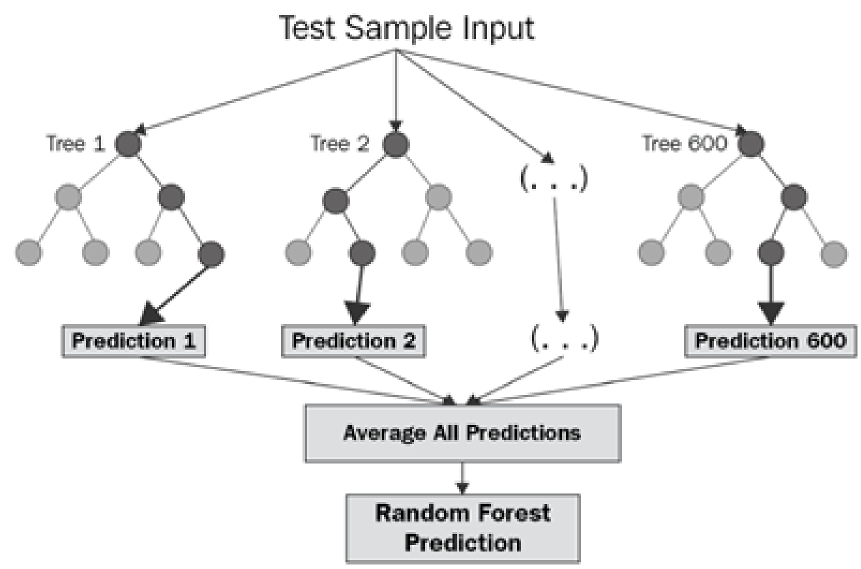

Figure 4. Random tree [30].

3.1.4 Decision tree

Decision trees are a non-linear predictive modeling tool used for classification and regression tasks. They mimic human decision-making processes by splitting data into branches at decision nodes, based on some decision criteria. Due to their simplicity and interpretability, decision trees are widely used in areas such as sentiment analysis, where they can help in analyzing text data for sentiment and intent, including in the contexts of financial stock market forecasts and analyzing the impact of press media on public sentiment [31].

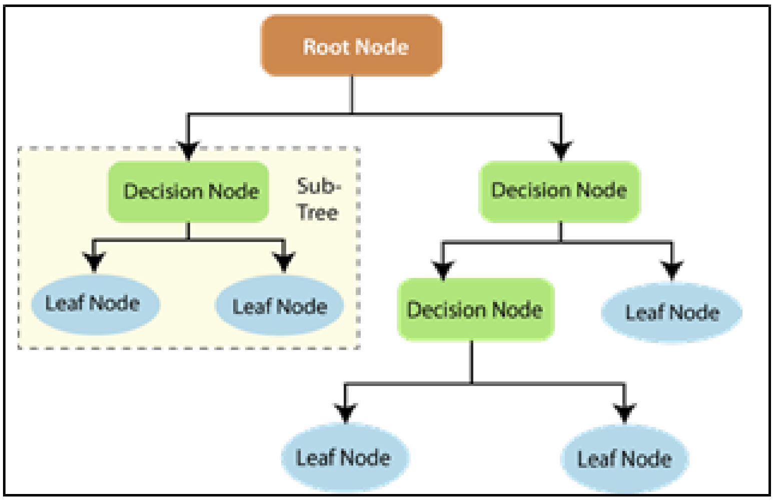

Figure 5. Decission tree [32].

3.1.5 Naive Bayes:

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. They are particularly suited for categorical input data and are extensively used in text classification, making them ideal for applications in sentiment analysis across various domains, such as news sentiment prediction, sentiment analysis in social media, and sentiment classification of financial tweets [33].

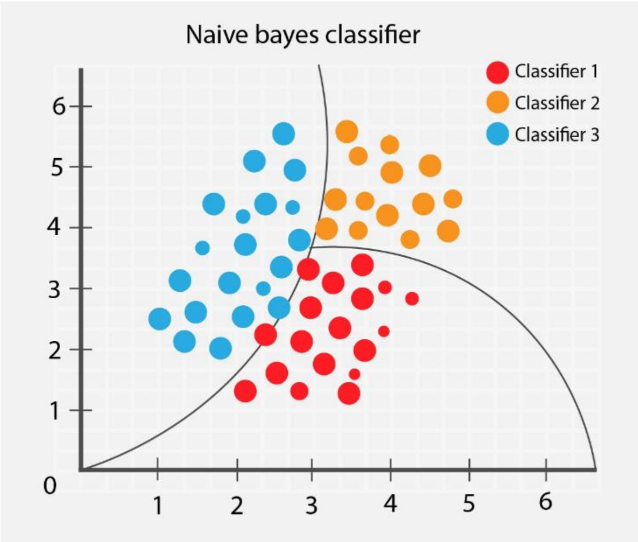

Figure 6. Naive Bayes [34].

3.2 Unsupervised Learning Algorithms

Unsupervised learning algorithms are employed in sentiment analysis to uncover hidden patterns or structures in unlabeled data. These algorithms are particularly useful in scenarios where labeled data is scarce or unavailable. Techniques such as clustering, dimensionality reduction, and topic modeling are common unsupervised learning methods used to analyze sentiments and opinions without prior knowledge of the data labels [35].

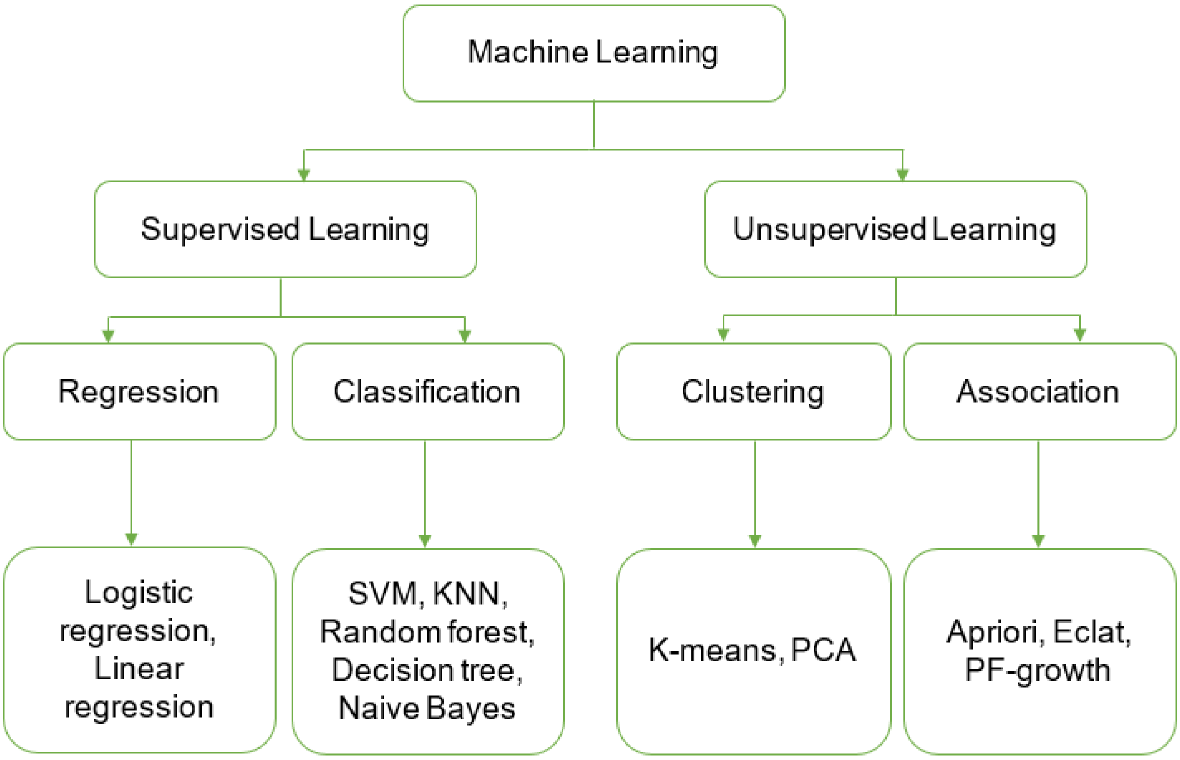

Figure 7. The structure of ML [36].

3.3 Hybrid-based Approach

The hybrid-based approach in sentiment analysis combines both supervised and unsupervised learning techniques to leverage the strengths of both methods. This approach often involves using unsupervised learning to explore and understand the underlying patterns in the data, followed by applying supervised learning algorithms to classify or predict sentiments. Hybrid models can provide more accurate and robust solutions by incorporating the advantages of both learning paradigms [37].

C. LITERATURE REVIEW

Hassan in 2022 [38] analyzed the effectiveness of machine learning algorithms for text classification, focusing on Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Logistic Regression (LR), Multinomial Naïve Bayes (MNB), and Random Forest (RF). Their study evaluated these algorithms on IMDB and Spam datasets, finding that Logistic Regression and SVM performed best on IMDB, while k-NN was most effective on Spam. k-NN achieved the highest accuracy (98.5%) for Spam, and Logistic Regression achieved 85.8% accuracy for IMDB. Further research into advanced algorithms and methods like hyperparameter tuning and ensembles was suggested to enhance text classification.

Sofia et al. [39] in 2023 introduced a machine learning model for detecting depression during the Covid-19 crisis, employing Python's capabilities and methods like Decision Tree, KNN, and Naive Bayes. Using a survey based on the Hamilton tool and expert advice, they aimed to identify depression. Results showed KNN had the highest accuracy (92.32%), while Decision Tree excelled in detecting depression quickly. This suggests machine learning, especially KNN for accuracy and Decision Tree for speed, could replace conventional methods for depression detection, emphasizing early intervention in mental health.

Akhtar et al. [40] in 2022 conducted a study on machine learning algorithms' effectiveness in predicting stock market prices. They compared Random Forest and Support Vector Machine (SVM) algorithms with traditional models, highlighting the importance of data preprocessing for accurate analysis. Achieving 80.3% prediction accuracy, the study demonstrated machine learning's potential in enhancing stock market predictions. It emphasized the need to integrate advanced algorithms and thorough data preprocessing for improved accuracy in stock market investments.

Van den Bulk et al. [41] in 2022 conducted a study on using machine learning (ML) algorithms to improve the efficiency of systematic reviews in food safety. They tested eight ML algorithms and their ensembles on datasets related to chemical hazards in cereals and leafy greens. The study found that an ensemble of Naive Bayes and Support Vector Machine algorithms reduced the workload on experts by decreasing the number of articles requiring manual review while maintaining 95% of the relevant literature. This research highlights the potential of ML to enhance systematic review processes without sacrificing quality.

Motz et al. [42] in 2022 focused on improving real-time sentiment analysis of tweets by combining various machine learning and text processing algorithms. They utilized Support-Vector Machine, Naive Bayes, Textblob, and the Lexicon Approach, achieving 68.29% accuracy. This method shows promise for fast and accurate sentiment analysis of live data streams, suggesting that blending algorithms can enhance sentiment detection in social media, providing a valuable solution for analyzing sentiments in dynamic data environments.

Sudhakar & Kaliyamurthie [43] in 2024 investigated the detection of fake news on social media during events like the COVID-19 pandemic, emphasizing its impact on public health. They analyzed over 1.3 million COVID-19-related tweets, testing various classifiers. SVM and Logistic Regression were most effective, achieving 98% and 95.2% accuracy, respectively. The study highlighted the necessity for robust algorithms to combat misinformation on social media and suggested continued research into advanced detection methods.

Leelawat et al. [44] in 2022 utilized machine learning to analyze tweets about tourism in Thailand's key destinations during the COVID-19 pandemic. They employed algorithms like Decision Tree, Random Forest, and SVM to understand sentiments and intentions toward tourism. SVM performed best for sentiment analysis, especially for Phuket-related tweets, achieving 77.4% accuracy. Random Forest excelled in analyzing intentions for Bangkok, with 95.4% accuracy. By analyzing common words in each sentiment and intention category, the study suggested strategies for promoting tourism, addressing COVID-19 concerns, and political instability. This research underscored the value of machine learning in extracting insights from social media for tourism stakeholders during crises.

Govindan & Balakrishnan [45] in 2022 investigated sarcasm detection in negative sentiment tweets related to COVID-19, focusing on hyperbolic expressions. They analyzed 6,600 tweets with specific hashtags, identifying and evaluating five hyperbolic features. Using machine learning algorithms such as Support Vector Machine, Random Forest, and Random Forest with Bagging, they found that the elongated word feature achieved the highest accuracy (78.74%) and F-score (71%), with intensifiers being the most crucial hyperbolic feature. This study contributed to understanding sarcasm detection in sentiment analysis, particularly in social media discussions during the COVID-19 pandemic, emphasizing the importance of hyperbolic expressions for accuracy improvement.

Pratama & Tjahyanto [46] in 2021 examined how fake accounts impacted sentiment analysis regarding COVID-19 in Indonesia. They introduced a new method that considered account authenticity, comparing Support Vector Machine (SVM) and Naïve-Bayes algorithms. Using tweets from the Indonesian Ministry of Health's Twitter account and Tweetbotornot tool to identify fake accounts, they found that fake accounts influenced sentiment analysis, with SVM showing better accuracy (80.6%) than Naïve-Bayes. Surprisingly, fake accounts tended to skew sentiment positively. The research emphasized the importance of considering account authenticity in sentiment analysis and highlighted SVM's effectiveness in handling such complexities.

Shahzad et al. [47] in 2023 conducted a study on Facebook users' emotional responses to research articles, utilizing machine learning techniques. They analyzed over 223,077 posts and found that Random Forest models performed best, achieving 86% accuracy for binary classifications and 66% for ternary classifications. The study emphasized the influence of research article titles on Facebook sentiment and highlighted the potential of sentiment analysis for understanding public engagement with scientific research. Its findings are essential for enhancing science communication and engagement strategies on social media, showcasing the integration of machine learning in sentiment analysis for a deeper understanding of public sentiment toward scientific research.

Dangsawang & Nuchitprasitchai [48] in 2024 examined customs fraud via social media, spurred by the growth of e-commerce. Unauthorized sellers on platforms like Twitter and Facebook caused substantial tax losses. They devised the SHIELD model, utilizing machine learning on a dataset of 2,373,570 social media records. Algorithms like Logistic Regression, Gated Recurrent Unit, and Long Short-Term Memory categorized posts into "Red Line" for potential fraud, "Green Line" for non-commercial goods, and "Inspect" for further review. LSTM proved most effective, with a 99.44% accuracy rate. This highlighted the efficacy of machine learning and NLP in identifying and mitigating customs fraud, promising enhanced economic security and regulatory enforcement.

Cahyana et al. [49] in 2022 presented a semi-supervised learning method using K-Nearest Neighbors (KNN) and Term Frequency-Inverse Document Frequency (TF-IDF) to automatically annotate hate speech on social media. This approach improved annotation accuracy from 57.25% to 59.68% by analyzing 13,169 comments, resulting in 11,235 annotated data points. Despite challenges with unannotated data, the study demonstrated the potential of combining KNN and TF-IDF to enhance hate speech detection, offering a scalable solution for large datasets and paving the way for future automated text analysis advancements.

Sarsam et al. [50] in 2022 introduced a novel machine learning mechanism for early disease recognition using sentiment analysis on Twitter, focusing on anemia. They utilized algorithms like k-means clustering and Latent Dirichlet Allocation (LDA) to analyze tweets, identifying both disease symptoms and patient emotions. Their approach was validated using various classifiers, with Sequential Minimal Optimization (SMO) achieving 98.96% prediction accuracy. The study highlighted fear and sadness as dominant emotions among anemic patients, suggesting the potential of social media for non-invasive disease diagnosis and intelligent health monitoring systems. This breakthrough method could be applied to other diseases, improving clinical decision support in resource-limited settings.

Gupta & Rattan [51] in 2023 presented a refined sentiment analysis framework for Twitter, using Support Vector Machine (SVM) with Particle Swarm Optimization (PSO) and an Effective Word Score (EFWS) heuristic. This method improved sentiment analysis accuracy by optimizing training sample selection and utilizing the EFWS heuristic. Tested on 1.6 million tweets, the model achieved approximately 80% accuracy, increasing to 85% with EFWS on a subset of 100K tweets, while also doubling the training speed. Their study underscored the potential of machine learning advancements to enhance sentiment analysis accuracy and efficiency, especially for analyzing large-scale social media data.

Lomotey et al.[52] in 2023 analyzed over 2.8 million tweets from people with HIV/AIDS during the COVID-19 pandemic, using machine learning techniques. Their study aimed to identify key concerns and sentiments, revealing 14 major themes including medical care accessibility and stigmatization. They found a mix of positive, negative, and neutral sentiments, with sentiment analysis achieving a high F1-score of 90%. The research highlighted the challenges faced by this group during the pandemic and demonstrated the potential of machine learning in extracting valuable insights from social media for informing healthcare policies and support mechanisms.

Hidayat et al. [53] in 2021 analyzed public sentiment on Rinca Island's development in Indonesia via Twitter. They used machine learning, including two Doc2Vec models, SVM, and Logistic Regression classifiers, to categorize sentiments. Results showed over 75% accuracy, indicating predominantly negative sentiment towards the project. The best accuracy (87%) and F1 score (81%) were with the PV-DBOW model combined with SVM. The study highlighted machine learning's role in understanding public opinion on social and environmental issues via social media, potentially impacting decision-making and policy regarding development projects.

Maulana et al. [54] in 2020 conducted research to boost sentiment analysis accuracy for movie reviews by integrating the Support Vector Machine (SVM) algorithm with Information Gain (IG) for feature selection. Their study showed that feature selection significantly affected SVM's performance in categorizing reviews into positive, negative, or neutral sentiments. By applying their method to the Cornell and Stanford datasets, they observed notable improvements in SVM's classification accuracy with the inclusion of IG. This enhancement suggested that combining SVM and IG offered a more effective approach for analyzing sentiment in movie reviews. The study's findings indicated potential for broader application and encouraged further exploration of this methodology across different review types and languages to advance sentiment analysis techniques in machine learning.

Macrohon et al. [55] in 2022 introduced a semi-supervised sentiment analysis method for assessing tweets from the 2022 Philippine Presidential Election, achieving an 84.83% accuracy rate. Their approach, focusing on English and Tagalog tweets, utilized Multinomial Naïve Bayes as the base classifier, enhanced by Natural Language Processing techniques. The method effectively categorized tweets into positive, neutral, and negative sentiments, providing insights into public opinion during the divisive election. This method surpassed previous accuracies in similar regional studies, highlighting the potential of semi-supervised learning in bilingual sentiment analysis for sociopolitical research and NLP advancements.

Maqbool et al. [56] in 2023 investigated how sentiment analysis from financial news, combined with historical stock data, could predict stock prices. They utilized algorithms like VADER, TextBlob, and FLAIR to assess sentiment scores' impact on stock market forecasts across various sectors. Introducing a machine learning model, specifically the MLP-Regressor, integrated with sentiment scores, achieved a notable prediction accuracy of up to 0.90. This study emphasized the crucial role of financial news sentiment in stock price fluctuations and showcased the potential of machine learning techniques to leverage sentiment for more accurate predictions in stock market forecasting.

Olabanjo et al. 2023 [57] in 2023 conducted a comprehensive study on applying sentiment analysis to understand public opinion during Nigeria's 2023 presidential election using Twitter data. Analyzing two million tweets, they employed machine learning models like LSTM, BERT, and LSVC to assess sentiments towards the leading candidates. Results indicated varying levels of public sentiment for each candidate, with BERT showing superior accuracy and precision. The study highlighted the importance of sentiment analysis in understanding political landscapes and provided insights into the election's potential outcome based on social media sentiment analysis.

Jadhav et al. [58] in 2023 investigated using machine learning to differentiate between abusive head trauma (AHT) and non-abusive head injuries in children under 5. They conducted a retrospective review, comparing AHT-diagnosed children to a control group. Applying a Random Forest (RF) ML algorithm, they combined sentiment and subjectivity analysis of physician notes with demographic data to predict AHT. Results showed significant improvement in AHT classification accuracy, achieving an AUC of 0.78 and overall accuracy of 84%. This study highlighted ML and natural language processing's potential in early AHT detection, offering clinicians a new tool for assessing possible abuse in pediatric patients.

Oyebode et al. [59] in 2020 evaluated mental health apps by analyzing user reviews using machine learning (ML) sentiment analysis and thematic analysis. They examined 104 apps and 88,125 reviews from Google Play and the App Store. Among five ML classifiers tested, Stochastic Gradient Descent performed the best (F1-score: 89.42%). Thematic analysis identified 21 negative and 29 positive themes, covering usability, content, ethical, customer support, and billing issues. This study highlighted the potential of combining quantitative and qualitative methods for a comprehensive evaluation of mental health apps, offering insights for developers to improve app effectiveness by addressing negative aspects and leveraging positive ones.

Nayak et al. [60] in 2023 introduced an innovative sentiment analysis approach, combining a modified Bayesian Boosting algorithm with weight-guided optimal feature selection. This method aimed to enhance sentiment classification by prioritizing informative features and optimizing the boosting process. The algorithm effectively addressed sentiment analysis complexities, notably improving classification performance by adjusting weights dynamically for misclassified instances. Compared to existing methods, it achieved superior results in accuracy, precision, recall, and F1 score, with an impressive accuracy of 98.49%. This research significantly advanced sentiment analysis by providing a more efficient model, surpassing traditional methods and paving the way for more precise and effective sentiment analysis tools.

Wadhe & Suratkar [61] in 2020 investigated machine learning algorithms for sentiment analysis of tourist place reviews. They utilized data from tourism review websites to explore CountVectorization and TFIDFVectorization for feature extraction, along with Naive Bayes, Support Vector Machine, and Random Forest for classification. Comparing their performance using various metrics, they found TFIDFVectorization enhanced accuracy, with TFIDFVectorization combined with Random Forest achieving the highest accuracy at 86%. This study emphasized the potential of machine learning in helping tourists make informed decisions by analyzing sentiments in online reviews.

Pavitha et al. [62] in 2022 introduced a system that combined movie recommendations with sentiment analysis on movie reviews using machine learning. They utilized Cosine Similarity for recommending movies based on user preferences and employed Naïve Bayes (NB) and Support Vector Machine (SVM) classifiers for sentiment analysis. The SVM classifier, with an accuracy of 98.63%, outperformed NB, providing more reliable classification of reviews as positive or negative. This innovative approach not only helped users discover movies matching their tastes but also offered insights into public sentiment, enabling more informed viewing choices.

Qorib et al. [63] in 2022 used sentiment analysis and machine learning to study COVID-19 vaccine hesitancy on Twitter. They employed various sentiment analysis methods and machine learning algorithms to categorize tweets. Results showed a decrease in vaccine hesitancy over time, with the TextBlob, TF-IDF, and LinearSVC combination performing best. This study highlighted the potential of sentiment analysis and machine learning in understanding public sentiment towards vaccination, offering insights for public health communication strategies.

Alsemaree et al. [64] in 2024 developed a pioneering method for analyzing customer sentiments in Arabic tweets regarding coffee products. Motivated by the intricacies of the Arabic language and the necessity for businesses in the competitive coffee market to grasp subtle customer opinions, the research employed advanced techniques such as TF-IDF and MRMR for feature extraction. Additionally, four machine learning algorithms were utilized, including k-nearest neighbor, support vector machine, decision tree, and random forest, combined with ensemble learning to enhance sentiment classification. With an accuracy surpassing 95.95%, the study introduced a highly effective ensemble model for Arabic sentiment analysis, providing valuable insights for product development and marketing strategies.

Bengesi et al. [65] in 2023 utilized machine learning to analyze more than 500,000 multilingual tweets concerning the monkeypox outbreak. They employed sentiment analysis tools VADER and TextBlob to categorize the tweets into positive, negative, and neutral sentiments. The objective was to provide policymakers and health authorities with insights into public perceptions of monkeypox, aiding in the formulation of effective health policies. Through rigorous data collection, preprocessing, and evaluation of 56 classification models using various machine learning algorithms, the study identified a highly accurate model combining TextBlob annotation, Lemmatization, CountVectorizer, and SVM, achieving around 93.48% accuracy. This research underscored the potential of sentiment analysis and machine learning in informing public health responses to emerging diseases.

Patel et al. [66] in 2023 examined various machine learning models, including BERT, to analyze sentiment from airline reviews. They found that BERT outperformed other models, achieving an 83% accuracy rate. This emphasized BERT's ability to accurately interpret customer sentiments. The study highlighted the potential of sentiment analysis, especially through advanced NLP techniques, to improve customer service in the airline industry. It contributed to the literature by demonstrating the practical advantages of using sophisticated NLP models like BERT to understand customer feedback and enhance business operations.

Vashishtha & Susan [67] In 2020 introduced an innovative approach to multimodal sentiment analysis, integrating supervised fuzzy rule-based systems. Their research showed significant advancements in understanding sentiments in social media video reviews by combining linguistic and acoustic modalities. Using a set of clever fuzzy rules, they classified sentiment based on confidence scores from SVM classifications of text and speech cues. This approach set a new standard by achieving an accuracy of 82.5%, surpassing existing models. Their work highlighted the potential of fuzzy rule-based systems in enhancing multimodal sentiment analysis effectiveness.

Table 1. Summary of studies conducted on sentiment analysis using ML Techniques.

#	Authors & Year	Datasets	Based Model	Accuracy	Advantages	Limitations
1	Hassan et al 2022 [38]	IMDB & Spam	Logistic Regression, SVM, k-NN	Spam: 98.5%, IMDB: 85.8%	Highly effective for spam detection	Computationally expensive, requires tuning of 'k' parameter
2	Sofia et al 2023 [39]	1,694 survey responses	Decision Tree, KNN, Naive Bayes	KNN: 92.32%	Innovative use of surveys, includes dynamic assessment	Limited to Covid-19 impact, relies on self-reported data
3	Akhtar et al 2022 [40]	Stock market price	Random Forest, SVM	80.3%	Utilizes advanced ML algorithms, focuses on data	Limited to historical data, may not

		dataset			preprocessing	account for unforeseen events
4	van den Bulk et al 2022 [41]	Chemical hazards in cereals and leafy greens	Eight ML algorithms including LR, SVM, NB, RF, AB, GB, LSTM, BERT	95%	Reduces manual review workload, maintains 95% relevance	Limited to food safety, requires existing data, may not generalize well
5	Motz et al 2022 [42]	Twitter API data stream	SVM, Naive Bayes, Textblob, Lexicon	68.29%	Effective for real-time sentiment analysis	Limited to Twitter data, accuracy suggests room for improvement
6	Sudhakar & Kaliyamurthie 2024 [43]	Covid-19 (1,375,592 tweets)	SVM, Logistic Regression	SVM: 98%, LR: 95.2%	Excellent for linear and non-linear data	Computationally intensive, dependent on kernel choice
7	Leelawat et al 2022 [44]	Over 150,000 English-language tweets related to tourism in Bangkok, Chiang Mai, and Phuket	Decision Tree, Random Forest, SVM	SVM: 77.4%, RF: 95.4%	High accuracy in sentiment analysis, actionable insights	Limited to English tweets, focuses on specific regions
8	Govindan & Balakrishnan 2022 [45]	6,600 pre-processed tweets with hashtags #Chinesevirus, #Kungflu, #COVID19, #Hantavirus, and #Coronavirus	SVM, Random Forest, Bagging	78.74%	Effectiveness of hyperbolic features in sarcasm detection	Focused solely on negative sentiments

9	Pratama & Tjahyanto 2021 [46]	Tweets from the Indonesian Ministry of Health's Twitter account	SVM, Naïve-Bayes, Tweetbotor not tool	SVM: 80.6%	Novel approach considering account authenticity	Limited to a single data source, may limit generalizability
10	Shahzad et al 2023 [47]	223,077 Facebook posts related to various scientific research domains	ML models including Random Forest, Decision Tree, K-NN, Logistic Regression, Naïve Bayes	Binary: 86%, Ternary: 66%	Large dataset, comprehensive sentiment analysis on Facebook	Complexity of data preprocessing, limited to Facebook
11	Dangsawang & Nuchitprasitchai 2024 [48]	2,373,570 records from Twitter and Facebook	Logistic Regression, GRU, LSTM	99.44%	High accuracy in detecting customs fraud, uses advanced ML techniques	Limited to English and Thai, extensive computational resources required
12	Cahyana et al 2022 [49]	13,169 comments from YouTube, with 2,370 initially labeled for training and 9,482 for testing after preprocessing.	K-NN, TF-IDF	59.68%	Reduced human annotation effort, efficient data handling	Complexity of annotation, may not generalize well outside tested domain
13	Sarsam et al 2022 [50]	1,738,759 English tweets from Dec 2019 to May 2020	K-means, LDA, Apriori, various classifiers including	98.96%	Non-invasive, cost-effective, uses widely available data	Limited to English tweets specific to anemia, extensive preprocessing

			SMO			needed
14	Gupta & Rattan 2023 [51]	1.6 million tweets	SVM with PSO and EFWS heuristic	Up to 85%	High accuracy, doubles training speed, integrates tweet subjectivity	Limited to Twitter data, requires computational resources for optimization
15	Lomotey et al 2023 [52]	Tweets related to HIV/AIDS	ML techniques: textual mining, thematic analysis, LDA, VADER	F1-score: 90%	Identified major themes, high F1-score with VADER	Limited to English tweets, potential bias in tweet selection
16	Hidayat et al 2021 [53]	Twitter data related to Rinca Island development	Doc2Vec, SVM, Logistic Regression	87%	Effective in analyzing public sentiment	Possible data imbalance, limited to Twitter data
17	Maulana et al 2020 [54]	Cornell and Stanford movie review datasets	SVM with Information Gain	86.62%	Improved accuracy in movie review sentiment analysis	Focus mainly on feature selection, limited comparison with other methods
18	Macrohon et al 2022 [55]	Tweets in English and Tagalog during the campaign period (Feb 8 to May 8, 2022) for the 2022 Philippine Presidential Election	Semi-supervised learning with Multinomial Naïve Bayes	84.83%	High accuracy in sentiment classification, handles large datasets with limited data	Focused only on top candidates, methodology may need adjustments for different contexts

19	Maqbool et al 2023 [56]	Historical stock data and financial news from various companies including Reliance, Tata Motors, Tata Steel, and HDFC	MLP-Regressor with sentiment analysis tools	Up to 90%	High accuracy, tested on diverse sectors	Complexity in preprocessing, reliance on accurate sentiment scoring, computational intensity
20	Olabanjo et al 2023 [57]	2 million tweets with 18 features from Twitter concerning the 2023 Presidential election in Nigeria	LSTM, BERT, LSVC	BERT: 94%	High accuracy in sentiment analysis, insights into public opinion	Limited by representativeness of Twitter users, may not capture full public opinion
21	Jadhav et al 2023 [58]	Pediatric patients < 5 years at a Level 1 Trauma Center, 587 non-AHT patients, 193 AHT patients	Random Forest	84%	Utilizes clinical documentation for early detection	Limited dataset size and scope, potential biases in clinical documentation, relies on initial docs
22	Oyebode et al 2020 [59]	88,125 user reviews from 104 mental health apps on Google Play and App Store	SVM, MNB, SGD, LR, RF	F1-score: 89.42%	Deep insights into user sentiment and app effectiveness	Relies on user reviews, may have subjective elements in theme categorization

23	Nayak et al 2023 [60]	Hotel reviews data from various hotels	Modified Bayesian Boosting algorithm	98.49%	Superior accuracy and sentiment analysis performance	Complexity in feature selection, limited applicability to similar datasets
24	Wadhe et al 2020 [61]	3,209 reviews from various tourism websites	Naive Bayes, SVM, Random Forest	86%	Improved accuracy with TFIDFVectorizatio n	TFIDF requires more execution time, scalability not discussed
25	Pavitha et al 2022 [62]	"tmdb_500 0_movies.cs v", "tbmd_500 0_credits.cs v" for movie recommen dation; "reviews.tx t" for sentiment analysis	Cosine Similarity, Naïve Bayes, SVM	SVM: 98.63%	Efficient movie recommendation and sentiment analysis	Limited to specific dataset and methodology, may not generalize to other contexts
26	Qorib et al 2022 [63]	42,796 tweets related to COVID-19 vaccination collected from Twitter using API search from September 26, 2021, to November 7, 2021	Azure ML, VADER, TextBlob, various ML algorithms	96.75%	Comprehensive approach, high accuracy in sentiment analysis	Limited to Twitter, may not represent broader public opinion, computational intensity mentioned
27	Alsemaree et al 2024 [64]	10,646 tweets about	Supervised learning with	Over 95.95%	High accuracy, suitable for Arabic text sentiment	Focused on Arabic text, complexity in

		various coffee products	ensemble methods		analysis	implementation
28	Bengesi et al 2023 [65]	500,000 multilingual tweets related to monkeypox from Twitter	VADER, TextBlob, CountVectorizer, SVM	93.48%	Large dataset, multilingual, variety of ML algorithms	Computational intensity, potential bias in sentiment tools
29	Patel et al 2023 [66]	Airline reviews dataset from Kaggle	BERT, various ML models	83%	BERT outperforms in understanding text nuances	Requires substantial computational resources, pre-processing needed
30	Vashishtha and Susan 2020 [67]	CMU MOSI	Supervised fuzzy rule-based system using SVM	82.5%	Innovative integration of fuzzy logic with sentiment analysis	Relies on predefined rules, complexity increases with more modalities

D. Discussion

Our comprehensive review analyzed a range of studies focusing on the integration of machine learning techniques in sentiment analysis. The studies surveyed employed various machine learning models on diverse datasets, such as Twitter data, movie reviews, stock market data, and user reviews on social media platforms. Key findings from the Literature Review Summary include high accuracy in specific applications, with models like SVM and LSTM often achieving high accuracy. For example, Hassan et al. in 2022 and Dangsawang & Nuchitprasitchai in 2024 demonstrated the effectiveness of advanced machine learning techniques in spam and fraud detection with accuracies above 98%.

Domain-specific challenges and solutions were also highlighted, as several studies adapted sentiment analysis models to specific contexts, achieving significant success. For example, van den Bulk et al. in 2022 analyzed chemical hazards in food with a 95% accuracy rate, while Leelawat et al. in 2022 tailored their models for tourism-related sentiment in Thailand.

Emerging methodologies such as ensemble learning and hybrid models showed promising results in enhancing sentiment analysis. For instance, Gupta & Rattan in 2023 employed a combination of SVM with heuristic methods to speed up training and improve accuracy in analyzing tweets.

Despite these advancements, many models faced limitations related to computational intensity and the generalizability of findings. Studies focusing on specific data types or languages often mentioned the need for extensive computational resources or highlighted challenges in applying the findings to broader datasets.

The practical applications of these models are vast, ranging from detecting public health sentiments during pandemics to analyzing financial market trends and consumer behavior on social media. The adaptability of machine learning techniques across varied contexts illustrates their potential to significantly impact real-world decision-making.

In conclusion, while advancements in machine learning have considerably improved the capabilities of sentiment analysis, the reviewed studies also point to an ongoing need for methodological innovation to address emerging challenges. The field is moving towards more sophisticated models that promise even greater accuracy and deeper insights, marking a dynamic evolution of sentiment analysis tools powered by machine learning.

E. Conclusion

The evolution of sentiment analysis, as fueled by advancements in machine learning, represents a significant leap forward in our ability to interpret and utilize the vast data generated in the digital age. The integration of machine learning techniques has not only enhanced the accuracy and depth of sentiment analysis but has also expanded its applications across diverse fields such as finance, health, and social media. This review highlights the critical contributions of sentiment analysis to understanding complex human emotions and opinions, offering invaluable insights for businesses, policymakers, and researchers. As we look towards the future, the potential for further innovation in sentiment analysis is vast, promising even more sophisticated tools for data analysis in our increasingly interconnected world. The challenges identified pave the way for future research, emphasizing the need for ongoing advancements in machine learning models and methodologies to address the complexities of sentiment analysis in a rapidly evolving digital landscape.

F. References

[1] S. M. Al-Ghuribi, S. A. Mohd Noah, and S. Tiun, “Unsupervised Semantic Approach of Aspect-Based Sentiment Analysis for Large-Scale User Reviews,” IEEE Access, vol. 8. Institute of Electrical and Electronics Engineers Inc., pp. 218592–218613, 2020. doi: 10.1109/ACCESS.2020.3042312.

[2] N. Braig, A. Benz, S. Voth, J. Breitenbach, and R. Buettner, “Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data,” IEEE Access, vol. 11. Institute of Electrical and Electronics Engineers Inc., pp. 14778–14803, 2023. doi: 10.1109/ACCESS.2023.3242234.

[3] M. Costola, O. Hinz, M. Nofer, and L. Pelizzon, “Machine learning sentiment analysis, COVID-19 news and stock market reactions,” Res Int Bus Finance, vol. 64, Jan. 2023, doi: 10.1016/j.ribaf.2023.101881.

[4] D. G. Dev, V. Bhatnagar, B. S. Bhati, M. Gupta, and A. Nanthaamornphong, "LSTMCNN: A hybrid machine learning model to unmask fake news," Heliyon, vol. 10, no. 3, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25244.

[5] H. He, G. Zhou, and S. Zhao, "Exploring E-Commerce Product Experience Based on Fusion Sentiment Analysis Method," IEEE Access, vol. 10, pp. 110248–110260, 2022, doi: 10.1109/ACCESS.2022.3214752.

[6] Y. Dai, J. Liu, J. Zhang, H. Fu, and Z. Xu, "Unsupervised Sentiment Analysis by Transferring Multi-source Knowledge," Cognit Comput, vol. 13, no. 5, pp. 1185–1197, Sep. 2021, doi: 10.1007/s12559-020-09792-8.

[7] Y. Mehta, A. Malhar, and R. Shankarmani, "Stock price prediction using machine learning and sentiment analysis," in 2021 2nd International Conference for Emerging Technology, INCET 2021, Institute of Electrical and Electronics Engineers Inc., May 2021. doi: 10.1109/INCET51464.2021.9456376.

[8] J. Kazmaier and J. H. van Vuuren, "The power of ensemble learning in sentiment analysis," Expert Syst Appl, vol. 187, Jan. 2021, doi: 10.1016/j.eswa.2021.115819.

[9] V. Mahalakshmi, Dr. P. Shenbagavalli, S. Raguvaran, Dr. V. Rajakumareswaran, and E. Sivaraman, "Twitter Sentiment Analysis Using Conditional Generative Adversarial Network," International Journal of Cognitive Computing in Engineering, Mar. 2024, doi: 10.1016/j.ijcce.2024.03.002.

[10] Q. Liu, W. S. Lee, M. Huang, and Q. Wu, "Synergy between stock prices and investor sentiment in social media," Borsa Istanbul Review, vol. 23, no. 1, pp. 76–92, Jan. 2023, doi: 10.1016/j.bir.2022.09.006.

[11] S. Lee, S. Ma, J. Meng, J. Zhuang, and T. Q. Peng, "Detecting Sentiment toward Emerging Infectious Diseases on Social Media: A Validity Evaluation of Dictionary-Based Sentiment Analysis," Int J Environ Res Public Health, vol. 19, no. 11, Jun. 2022, doi: 10.3390/ijerph19116759.

[12] A. Yadav, C. K. Jha, A. Sharan, and V. Vaish, "Sentiment analysis of financial news using unsupervised approach," in Procedia Computer Science, Elsevier B.V., 2020, pp. 589–598. doi: 10.1016/j.procs.2020.03.325.

[13] K. Ueda et al., "SSCDV: Social media document embedding with sentiment and topics for financial market forecasting," Expert Syst Appl, vol. 245, Jul. 2023, doi: 10.1016/j.eswa.2023.122988.

[14] A. M. Shah, X. Yan, A. Qayyum, R. A. Naqvi, and S. J. Shah, "Mining topic and sentiment dynamics in physician rating websites during the early wave of the COVID-19 pandemic: Machine learning approach," Int J Med Inform, vol. 149, May 2021, doi: 10.1016/j.ijmedinf.2021.104434.

[15] S. Hinduja, M. Afrin, S. Mistry, and A. Krishna, "Machine learning-based proactive social-sensor service for mental health monitoring using twitter data," International Journal of Information Management Data Insights, vol. 2, no. 2, Nov. 2022, doi: 10.1016/j.jjimei.2022.100113.

[16] Y. Li, Z. Lin, and S. Xiao, "Using social media big data for tourist demand forecasting: A new machine learning analytical approach," Journal of Digital Economy, vol. 1, no. 1, pp. 32–43, Jun. 2022, doi: 10.1016/j.jdec.2022.08.006.

[17] H. Zhang et al., "Analysis of public opinion on food safety in Greater China with big data and machine learning," Curr Res Food Sci, vol. 6, Jan. 2023, doi: 10.1016/j.crfs.2023.100468.

[18] M. Demircan, A. Seller, F. Abut, and M. F. Akay, "Developing Turkish sentiment analysis models using machine learning and e-commerce data," International Journal of Cognitive Computing in Engineering, vol. 2, pp. 202–207, Jun. 2021, doi: 10.1016/j.ijcce.2021.11.003.

[19] K. Rybinski, "Content still matters. A machine learning model for predicting news longevity from textual and context features," Inf Process Manag, vol. 60, no. 4, Jul. 2023, doi: 10.1016/j.ipm.2023.103398.

[20] D. Marutho, Muljono, S. Rustad, and Purwanto, "Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism," Journal of Open Innovation: Technology, Market, and Complexity, vol. 10, no. 1, Mar. 2024, doi: 10.1016/j.joitmc.2024.100211.

[21] L. Kelebercová and N. C. Zozuk, "Sentiment classification of annual reports in Slovak language using symbolic, subsymbolic and statistic methods," in Procedia Computer Science, Elsevier B.V., 2023, pp. 3508–3516. doi: 10.1016/j.procs.2023.10.346.

[22] N. Braig, A. Benz, S. Voth, J. Breitenbach, and R. Buettner, "Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data," IEEE Access, vol. 11. Institute of Electrical and Electronics Engineers Inc., pp. 14778–14803, 2023. doi: 10.1109/ACCESS.2023.3242234.

[23] A. J. Najafabadi, A. Skryzhadlovska, and O. F. Valilai, "Agile Product Development by Prediction of Consumers' Behaviour; using Neurobehavioral and Social Media Sentiment Analysis Approaches," Procedia Comput Sci, vol. 232, pp. 1683–1693, 2024, doi: 10.1016/j.procs.2024.01.166.

[24] M. Viggiato, D. Lin, A. Hindle, and C. P. Bezemer, "What Causes Wrong Sentiment Classifications of Game Reviews," IEEE Trans Games, 2021, doi: 10.1109/TG.2021.3072545.

[25] P. Hajek and R. Henriques, "Predicting M&A targets using news sentiment and topic detection," Technol Forecast Soc Change, vol. 201, Apr. 2024, doi: 10.1016/j.techfore.2024.123270.

[26] M. Mohammadi et al., "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems," Journal of Network and Computer Applications, vol. 178. Academic Press, Mar. 15, 2021. doi: 10.1016/j.jnca.2021.102983.

[27] S. M. Sarsam, A. I. Alzahrani, and H. Al-Samarraie, "Early-stage pregnancy recognition on microblogs: Machine learning and lexicon-based approaches," Heliyon, vol. 9, no. 9, Sep. 2023, doi: 10.1016/j.heliyon.2023.e20132.

[28] B. T. Chicho, A. M. Abdulazeez, D. Q. Zeebaree, and D. A. Zebari, "Machine Learning Classifiers Based Classification For IRIS Recognition," Qubahan Academic Journal, vol. 1, no. 2, pp. 106–118, Mar. 2021, doi: 10.48161/qaj.v1n2a48.

[29] J. Navarro, J. U. Piña, F. M. Mas, and R. Lahoz-Beltra, "Press media impact of the Cumbre Vieja volcano activity in the island of La Palma (Canary Islands): A machine learning and sentiment analysis of the news published during the

volcanic eruption of 2021," International Journal of Disaster Risk Reduction, vol. 91, Jun. 2023, doi: 10.1016/j.ijdrr.2023.103694.
[30] S. V. Amanoul, A. M. Abdulazeez, D. Q. Zeebare, and F. Y. H. Ahmed, "Intrusion Detection Systems Based on Machine Learning Algorithms," in 2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Jun. 2021, pp. 282–287. doi: 10.1109/I2CACIS52118.2021.9495897.
[31] J. M. Sangeetha and K. J. Alfia, "Financial stock market forecast using evaluated linear regression based machine learning technique," Measurement: Sensors, vol. 31, Feb. 2023, doi: 10.1016/j.measen.2023.100950.
[32] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
[33] M. Wilksch and O. Abramova, "PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets," International Journal of Information Management Data Insights, vol. 3, no. 1, Apr. 2023, doi: 10.1016/j.jjimei.2023.100171.
[34] J. Antas, R. R. Silva, and J. Bernardino, "Assessment of SQL and NoSQL Systems to Store and Mine COVID-19 Data," Computers, vol. 11, no. 2, Feb. 2022, doi: 10.3390/computers11020029.
[35] K. Rajalingam, N. Levin, O. Marques, J. Grichnik, A. Lin, and W. S. Chen, "Treatment options and emotional well-being in patients with rosacea: An unsupervised machine learning analysis of over 200,000 posts," JAAD Int, vol. 13, pp. 172–178, Dec. 2023, doi: 10.1016/j.jdin.2023.07.012.
[36] A. M. Iddrisu, S. Mensah, F. Boafo, G. R. Yeluripati, and P. Kudjo, "A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector," International Journal of Information Management Data Insights, vol. 3, no. 2, Nov. 2023, doi: 10.1016/j.jjimei.2023.100180.
[37] M. S. Satu et al., "TClustVID: A novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets," Knowl Based Syst, vol. 226, Aug. 2021, doi: 10.1016/j.knosys.2021.107126.
[38] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," Sustainable Operations and Computers, vol. 3, pp. 238–248, Jan. 2022, doi: 10.1016/j.susoc.2022.03.001.
[39] Sofia, A. Malik, M. Shabaz, and E. Asenso, "Machine learning based model for detecting depression during Covid-19 crisis," Sci Afr, vol. 20, Jul. 2023, doi: 10.1016/j.sciaf.2023.e01716.
[40] M. M. Akhtar, A. S. Zamani, S. Khan, A. S. A. Shatat, S. Dilshad, and F. Samdani, "Stock market prediction based on statistical data using machine learning algorithms," J King Saud Univ Sci, vol. 34, no. 4, Jun. 2022, doi: 10.1016/j.jksus.2022.101940.
[41] L. M. van den Bulk, Y. Bouzembrak, A. Gavai, N. Liu, L. J. van den Heuvel, and H. J. P. Marvin, "Automatic classification of literature in systematic reviews on food safety using machine learning," Curr Res Food Sci, vol. 5, pp. 84–95, Jan. 2022, doi: 10.1016/j.crfs.2021.12.010.
[42] A. Motz, E. Ranta, A. S. Calderon, Q. Adam, F. Alzhouri, and D. Ebrahimi, "Live Sentiment Analysis Using Multiple Machine Learning and Text Processing

Algorithms," in Procedia Computer Science, Elsevier B.V., 2022, pp. 165–172. doi: 10.1016/j.procs.2022.07.023.
[43] M. Sudhakar and K. P. Kaliyamurthie, "Detection of fake news from social media using support vector machine learning algorithms," Measurement: Sensors, vol. 32, p. 101028, Apr. 2024, doi: 10.1016/j.measen.2024.101028.
[44] N. Leelawat et al., "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," Heliyon, vol. 8, no. 10, Oct. 2022, doi: 10.1016/j.heliyon.2022.e10894.
[45] V. Govindan and V. Balakrishnan, "A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 8, pp. 5110–5120, Sep. 2022, doi: 10.1016/j.jksuci.2022.01.008.
[46] R. P. Pratama and A. Tjahyanto, "The influence of fake accounts on sentiment analysis related to COVID-19 in Indonesia," in Procedia Computer Science, Elsevier B.V., 2021, pp. 143–150. doi: 10.1016/j.procs.2021.12.128.
[47] M. Shahzad, C. Freeman, M. Rahimi, and H. Alhoori, "Predicting Facebook sentiments towards research," Natural Language Processing Journal, vol. 3, p. 100010, Jun. 2023, doi: 10.1016/j.nlp.2023.100010.
[48] B. Dangsawang and S. Nuchitprasitchai, "A machine learning approach for detecting customs fraud through unstructured data analysis in social media," Decision Analytics Journal, vol. 10, Mar. 2024, doi: 10.1016/j.dajour.2024.100408.
[49] N. Heri Cahyana, S. Saifullah, Y. Fauziah, A. Sasmito Aribowo, R. Drezewski, and P. Nasional Veteran Yogyakarta Yogyakarta, "Semi-supervised Text Annotation for Hate Speech Detection using K-Nearest Neighbors and Term Frequency-Inverse Document Frequency," 2022. [Online]. Available: www.ijacsa.thesai.org
[50] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, and A. S. Shibghatullah, "A non-invasive machine learning mechanism for early disease recognition on Twitter: The case of anemia," Artif Intell Med, vol. 134, Dec. 2022, doi: 10.1016/j.artmed.2022.102428.
[51] V. Gupta and P. Rattan, "Improving Twitter Sentiment Analysis Efficiency with SVM-PSO Classification and EFWS Heuristic," in Procedia Computer Science, Elsevier B.V., 2023, pp. 698–715. doi: 10.1016/j.procs.2023.12.125.
[52] R. K. Lomotey, S. Kumi, M. Hilton, R. Orji, and R. Deters, "Using Machine Learning to Establish the Concerns of Persons With HIV/AIDS During the COVID-19 Pandemic From Their Tweets," IEEE Access, vol. 11, pp. 37570–37601, 2023, doi: 10.1109/ACCESS.2023.3267050.
[53] T. H. J. Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," in Procedia Computer Science, Elsevier B.V., 2021, pp. 660–667. doi: 10.1016/j.procs.2021.12.187.
[54] R. Maulana, P. A. Rahayuningsih, W. Irmayani, D. Saputra, and W. E. Jayanti, "Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain," in Journal of Physics: Conference

Series, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1742-6596/1641/1/012060.
[55] J. J. E. Macrohon, C. N. Villavicencio, X. A. Inbaraj, and J. H. Jeng, "A Semi-Supervised Approach to Sentiment Analysis of Tweets during the 2022 Philippine Presidential Election," Information (Switzerland), vol. 13, no. 10, Oct. 2022, doi: 10.3390/info13100484.
[56] J. Maqbool, P. Aggarwal, R. Kaur, A. Mittal, and I. A. Ganaie, "Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach," in Procedia Computer Science, Elsevier B.V., 2022, pp. 1067–1078. doi: 10.1016/j.procs.2023.01.086.
[57] O. Olabanjo et al., "From Twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election," Heliyon, vol. 9, no. 5, May 2023, doi: 10.1016/j.heliyon.2023.e16085.
[58] P. Jadhav et al., "Application of a Machine Learning Algorithm in Prediction of Abusive Head Trauma in Children," J Pediatr Surg, vol. 59, no. 1, pp. 80–85, Jan. 2023, doi: 10.1016/j.jpedsurg.2023.09.027.
[59] O. Oyebode, F. Alqahtani, and R. Orji, "Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews," IEEE Access, vol. 8, pp. 111141–111158, 2020, doi: 10.1109/ACCESS.2020.3002176.
[60] S. Nayak, Savita, and Y. K. Sharma, "A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis," Decision Analytics Journal, vol. 8, Sep. 2023, doi: 10.1016/j.dajour.2023.100289.
[61] A. A. Wadhe and S. S. Suratkar, "Tourist Place Reviews Sentiment Classification Using Machine Learning Techniques," in 2020 International Conference on Industry 4.0 Technology, I4Tech 2020, Institute of Electrical and Electronics Engineers Inc., Feb. 2020, pp. 1–6. doi: 10.1109/I4Tech48345.2020.9102673.
[62] N. Pavitha et al., "Movie recommendation and sentiment analysis using machine learning," Global Transitions Proceedings, vol. 3, no. 1, pp. 279–284, Jun. 2022, doi: 10.1016/j.gltp.2022.03.012.
[63] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," Expert Syst Appl, vol. 212, Feb. 2023, doi: 10.1016/j.eswa.2022.118715.
[64] O. Alsemaree, A. S. Alam, S. Gill, and S. Uhlig, "Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions," Heliyon, p. e27863, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27863.
[65] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," IEEE Access, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.
[66] A. Patel, P. Oza, and S. Agrawal, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," in Procedia Computer Science, Elsevier B.V., 2023, pp. 2459–2467. doi: 10.1016/j.procs.2023.01.221.

[67] S. Vashishtha and S. Susan, "Inferring Sentiments from Supervised Classification of Text and Speech cues using Fuzzy Rules," in Procedia Computer Science, Elsevier B.V., 2020, pp. 1370–1379. doi: 10.1016/j.procs.2020.03.348.