Kashmir Journal of Science

https://kjs.org.pk **ISSN: 2958-7832**

Kashmir Journal of Science (2025), 4(3): 37-55; DOI: https://doi.org/10.63147/krjs.v4i3.81

Review Paper

Review of Social Media Sentiment Analysis Methods

Fazal Tariq[1,2], Muhammad Tufail[1], Taj Rehman[2]
[1]Department of Computer Science, GPGC, Nowshera, Pakistan
[2]Qutaba University of Science and Technology, Peshawar, Pakistan

*Corresponding Author: profazaltariq@gmail.com
http://orcid.org/0009-0008-8516-2442

ARTICLE INFO

Article history:

Received: 10 Nov 2024

Revised: 07 Feb 2025

Accepted: 17 Feb 2025

Available online: 30 Sep 2025

Keywords:

Social Media Channels (SMCs), Sentiment Analysis (SA), Machine Learning (ML), Deep Learning (DL), Big Data Analysis (BDA)

Abstract

Social media channels (SMCs) have become vital for communication, networking, and collaboration, facilitating rapid information sharing and business advertising. SMCs gather public opinions and emotions through user-generated content, essential for sentiment analysis. Companies and researchers analyze this data to understand consumer sentiment, trends, and preferences, enabling customized marketing strategies and enhanced customer engagement. This survey examines sentiment analysis and opinion mining techniques in social media using research from 2012 to 2023, covering lexicon-based, machine learning, and deep learning models. The study reviews recent research, demonstrating the application of these methods to assess public opinions and sentiments. It provides insights into current and future sentiment analysis techniques, offering recommendations for researchers, practitioners, and policymakers. Furthermore, the review explores the future potential of real-time data processing by integrating diverse data types, highlighting the value of social media data in refining sentiment analysis.

Introduction

Social media has changed the way people and organization keep in touch. It is a library of human minds in text, images and videos. The sheer volume of user generated content (Chen & Lin, 2014) poses a challenge and an opportunity to uncover in this corpus the sentiment expressed. A sentiment analysis tool (Wu & Li, 2016) is simply a computer software that can detect and extract opinions from written material as subjective information for the use reference in measuring public opinion, monitoring brand reputation so on to anticipate market trend

(Sánchez-Núñez et al., 2020). Sentiment analysis of textual data has become much more challenging with increase in the volume and velocity of text, but at the same time it also opens up many opportunities (Amangeldi et al., 2020; Barik & Misra, 2024). This adds in additional convoluted layers of the very colloquial or conversational language that traditional text analysis approaches were not built to manage but has become pervasive on these platforms. It is also needed to have update the sentiment analysis model adaption with new internet slang, emoji and social media idols.

This review article provides a view on Social Media Sentiment Analysis (SMSA). Various machine learning and deep learning techniques (Abulwafa, 2022; Alotaibi, 2023; Garg et al., 2020; Khan et al., 2022; Rustam et al., 2021; T.k. et al., 2021; Wu & Lode, 2020) are evolved from traditional lexicon-based approaches (Barik & Misra, 2024; Khan et al., 2023; van den Broek-Altenburg & Atherly, 2019), which this article also covers in a practical sense. This study will also mention some solutions for these challenges, noise and sarcasm in context of social media data. It will also explore several uses of sentiment analysis such as in marketing, politics, crisis management and public health. It works on adding value to sentiment analysis in social media by integrating the current studies and highlighting the future directions. We will discuss review of existing methodologies, major challenges and limitations of the previous works faced which leads to categorization mentioned in four broad categories.

Iglesias and Moreno (Iglesias & Moreno, 2019) highlighted the significance of sentiment analysis for social media platforms such as Twitter and Facebook in their work "Sentiment Analysis for Social Media." This has spurred the development of advanced, rapid sentiment analysis techniques for these platforms. Their paper delves into methods for measuring sentiment in social media data, addressing the challenges of extraction. With social media now a key outlet for public expression, sentiment analysis is vital for gauging customer feelings and understanding market and social dynamics. The authors introduce their dataset, comprising Twitter and Facebook posts, and discuss the difficulties of sentiment analysis due to the heterogeneity of social media data, including emojis, images, and videos. Posts are harder to interpret without context, and irony and humor can be misinterpreted by tools lacking multilingual support. Additionally, social media data is vast and streaming, necessitating real-time processing to handle its volume and velocity.

The article argues that social media texts, being informal and varied, are complex for sentiment analysis, requiring specialized approaches. It addresses multiple issues beyond cross-platform sentiment analysis, focusing on challenges with Twitter data. High-dimensional data increases computational difficulty and overfitting risk, and the informal nature of social media text complicates analysis. Sentiment analysis often uses part of the dataset, but noise, dimensionality, data domain, and varying training/testing data sizes can introduce errors. Dimension reduction eliminates unnecessary features, creating a space with fewer dimensions. "Sentiment Analysis on Social Media Tweets using Dimensionality Reduction and Natural Language Processing" (Omuya et al., 2023) explains how this aids understanding platforms like Twitter. The study analyzed over 1.6 million Sentiment140 tweets categorized by emoticons into positive, negative, and neutral. It utilized models beyond Naive Bayes, SVMs, and K-nearest neighbors, including those for dimensionality reduction, natural language processing, and part-of-speech tagging. This model significantly improved the accuracy with which sentiment was evaluated.

Social media serves as an emotional repository, where user-generated content is often consulted before online purchases. For smaller review scales, sentiment analysis tools can expedite this process. Utilizing machine learning and natural language processing technologies (Greene, 2018), BI (Behavior Insights) teams can discern consumers' motivations for engaging or not engaging with specific topics or products. Sentiment analysis, applicable in fields like Political Science, Sociology, and Psychology, detects trends, biases, and influences behind opinions. It evaluates emotionality in various platforms' opinions, including audio, written, and visual formats.

Automated sentiment analysis (Halawani et al., 2023) examines large web-based news articles, discussion groups, review blogs, and social media posts using machine learning, statistics, and natural language processing. Our research presents a framework incorporating dimensionality reduction, NLP techniques, and POS GUIDANCE FOR STS. Figure 1 illustrates a "Sentiment Analysis Model Source" for extracting relevant parts of speech to refine sentiment understanding and reduce input data size. Online assessment tools aggregate information from multiple sources, including social media, for improved evaluations. Sentiment analysis employs NB, SVM, and K-nearest neighbor models, with PCA and IG for optimal feature selection.

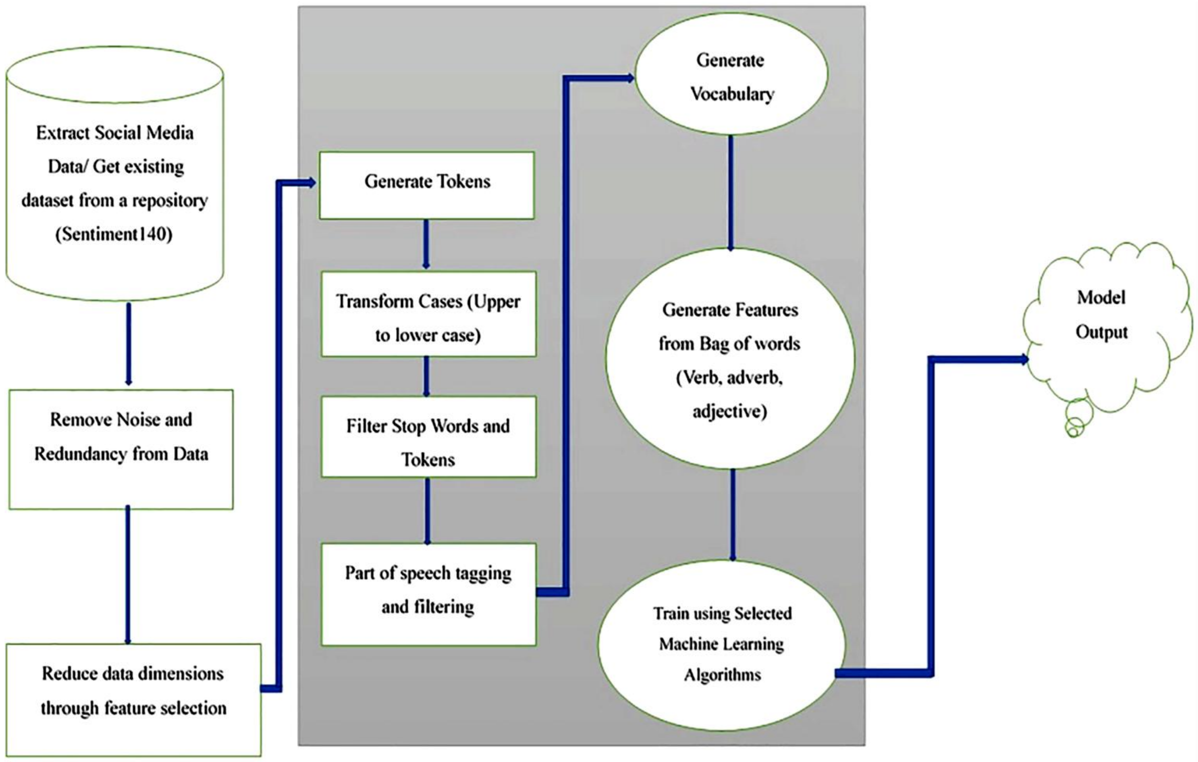

Fig.1 Sentiment analysis model

Review of Social Media Sentiment Analysis Methods

This model underpins a sentiment analysis algorithm, processing input (e.g., social media data) in real-time to quickly detect sentiment polarity in smaller datasets while maintaining precision and recall. Preprocessing social media and other data sources precedes model building, with data often sourced via APIs from social networks or repositories. The second step involves extracting features using PCA and information gain, followed by sentiment analysis through different parts of speech. We provide an end-to-end example, from training tokenization to word-bag creation. The algorithm, using NB/SVM/K-nearest neighbor, builds a classifier for text sentiment, evaluated during the testing phase before classifying new datasets or performing sentiment analysis. Table-1 presents metrics (accuracy, recall and precision) computed to evaluate the performance of proposed model. This article outlines sentiment analysis on social media and other datasets using SVM, K-nearest neighbor and Naïve Baise (NB) ML algorithms.

Table-1: Comparison of Machine Learning Techniques

Aspects/Techniques	KNN	SVM	NB
Model's Accuracy with 30% (test samples) and 70% (train samples)			
Proposed model for sentiment analysis	98.20%	90.15%	99.00%
Ensemble model build using selected features	79.25%	78.10%	76.50%
Sentiment analysis model build using short text	84.00%	81.45%	63.24%
Results computed using proposed sentiment analysis model with 30% (test samples) and 70% (train samples)			
Accuracy	98.20%	90.15 %	99.00 %
Precision	99.00	100.00	96.45
Recall	100.00	90.13	100.00
F-measure	97.60	91.90	98.00
Results computed using proposed sentiment analysis model and 10-fold cross validation			
Accuracy	100.00%	91.13%	99.00%
Precision	99.00	100.00	98.00
Recall	100.00	91.13	100.00
F-measure	98.00	93.90	98.00

The proposed model was benchmarked against top models for sentiment analysis through a performance evaluation. The model's performance was compared to other contemporary models used in sentiment analysis experiments. The authors highlighted the effectiveness of sentiment analysis models using different parts-of-speech, data sets, and dimension reduction processes, which improved classification by reducing noise and uncertainty. However, the limitations of this ML modeling suggest the need for further study with other model-development techniques. Rodriguez-Ibanez and Antonio's paper on “A review of sentiment analysis from social media platforms” (Rodríguez-Ibánez et al., 2023) underscores its growing importance in marketing, politics, and public opinion monitoring. Unlike traditional surveys or polls, public sentiment analysis on such platforms offers significant advantages for monitoring purposes.

Yin and Jin's survey “A survey of sentiment analysis on social media” (Ying et al., 2020) discusses the implementation of sentiment analysis in markets, politics, and public opinion monitoring systems (POMS) using large-scale unstructured text data from social media. It addresses the challenges and opportunities of using social media big data for sentiment analysis across various disciplines. Yard's article “Sentiment Analysis on Social Media” (Yadav, 2023) reviews existing work on sentiment analysis, focusing on multilingual datasets and the use of deep learning architectures, including CNNs and RNNs, to address these challenges. Over a hundred articles were reviewed, but it is impractical to discuss all related research in detail here. A summary of related studies is presented in Table 2.

Table-2: Summary of the relevant SMSA papers

SNO	Date of Pub	Authors	Title	Methodology	Achievements
1	2009	Go et al.	Twitter Sentiment Classification using Distant Supervision (Go et al., 2009).	Naive Bayes with distant supervision	Baseline for Twitter sentiment analysis
2	2013	Socher et al.	Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.(Socher et al., 2013)	Recursive Neural Networks (RNNs)	Enhanced hierarchical representation of sentences
3	2018	Zhang and Wang	Deep Learning for Sentiment Analysis: A Survey (Zhang et al., 2018)	CNNs & LSTMs	Outperformed traditional machine learning models
4	2023	Zhang et al.	Aspect-Based Sentiment Analysis: A Survey (Zhang et al., 2023)	Aspect-Based Sentiment Analysis (ABSA)	Improved granularity in sentiment classification
5	2019	Devlin et al.	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., 2019)	BERT (transformer-based model)	State-of-the-art performance in sentiment analysis
6	2019	Yang et al.	XLNet: Generalized Autoregressive Pretraining for Language Understanding (Yang et al., 2019)	XLNet (transformer-based model)	Improved performance over BERT in several tasks
7	2023	Hota H. S. et al.	Integration of deep learning techniques	Survey of deep learning methods	Overview of advancements in

Review of Social Media Sentiment Analysis Methods

SNO	Date of Pub	Authors	Title	Methodology	Achievements
			for sentiment and emotion analysis of social media data (Hota et al., 2023)		deep learning for sentiment analysis
8	2019	Howard and Ruder	Universal Language Model Fine-tuning for Text Classification (Howard & Ruder, 2018)	ULMFiT (transfer learning)	State-of-the-art results with fewer training examples
9	2016	Tang et al.	Aspect Level Sentiment Classification with Deep Memory Network (Tang et al., 2016)	Deep memory network	Effective for aspect-level sentiment analysis
10	2018	Peters et al.	Deep Contextualized Word Representations (Peters et al., 2018)	ELMo (deep contextualized embeddings)	Improved performance in various NLP tasks
11	2022	Ismet H et al.	Aspect based Sentiment of product review using memory network (Ismet et al., 2022)	Aspect based model	Better product review in coordination with memory network.
12	2014	Kim	Convolutional Neural Networks for Sentence Classification (Kim, 2014)	CNNs	Effective for short text classification
13	2014	dos Santos and Gatti	Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts (Dos Santos & Gatti, 2014)	CNNs	High accuracy for short text sentiment analysis
14	2014	Kalchbrenner et al.	A Convolutional Neural Network for Modelling Sentences(Kalchbrenner et al., 2014)	Dynamic CNNs	Captured sentence structure more effectively
15	2017	Vaswani et al.	Attention is All You Need (Vaswani et al., 2017)	Transformer Architecture	Foundation for BERT, GPT, and other models
16	2019	Devlin et al.	Multimodal Machine Learning: A Survey	Multimodal Sentiment Analysis	Comprehensive survey of

SNO	Date of Pub	Authors	Title	Methodology	Achievements
			and Taxonomy (Baltrusaitis et al., 2019)		multimodal approaches
17	2017	Miyato et al.	Adversarial Training Methods for Semi-Supervised Text Classification (Miyato et al., 2017)	Adversarial training	Improved robustness of text classifiers
18	2020	Brown et al.	Language Models are Few-Shot Learners (Brown et al., 2020)	GPT-3 (Generative Pre-trained Transformer 3)	Few-shot learning capability
19	2020	He et al.	BERT-MK: Integrating Graph and Tree Structures for Aspect-based Sentiment Analysis(Amangeldi et al., 2020)	BERT with graph and tree structures	Improved aspect-based sentiment analysis
23	2018	Xia et al.	Multi-grain Attention Network for Aspect-Level Sentiment Classification(Fan et al., 2018)	Multi-grain attention network	Enhanced aspect-level sentiment classification
24	2020	Ibanez M. et al.	Sentiment analysis applied to analyze society emotions in two different context of social media data(Ibañez et al., 2020)	Auto-encoder and deep learning are applied in conjunction with NLP text analysis techniques.	High performance on social media sentiment tasks
25	2019	Liu et al.	A Survey of Sentiment Analysis in Social Media (Yue et al., 2019)	Comprehensive survey	Detailed overview of methods and applications
26	2019	Sham et al.	Sentiment Analysis of Twitter data: A hybrid approach (Srivastava et al., 2019)	Hybrid deep learning	Enhanced performance by combining multiple approaches
27	2023	Sun et al.	Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis (Jain et al., 2023)	BERT-DCNN model	High accuracy and efficiency

Review of Social Media Sentiment Analysis Methods

SNO	Date of Pub	Authors	Title	Methodology	Achievements
28	2020	Almeida et al.	Explainable sentiment analysis application for social media crisis management in retail (Cirqueira et al., 2020)	Real-time sentiment analysis	Effective for managing crises with timely insights
29	2024	Başarslan & Kayaalp	Sentiment analysis using deep ensemble learning model (Başarslan & Kayaalp, 2024)	Ensemble methods	Improved accuracy and robustness with ensemble learning
30	2024	Amangeldi et al.	Understanding Environmental Posts: Sentiment and Emotion Analysis of Social Media Data (Amangeldi et al., 2024)	Using the Pointwise Mutual Information (PMI) algorithm	identifying patterns and trends in opinions related to the environment

Gaps and Challenges Identified

The work "Sentiment Analysis in Social Media" claims that the informal and noisy nature of social media texts complicates processing, necessitating specialized sentiment analysis algorithms. This paper aims to address key challenges in sentiment analysis on social media, particularly Twitter. High-dimensional data complicates computation and increases the risk of overfitting. Understanding sentiment in informal social media text is challenging due to noise, dimensionality, and data domain limitations. Dimension reduction can eliminate useless features and noise. Feature extraction transforms data into a low-dimensional space, whereas feature selection reduces dimensionality by removing unimportant features.

"A Review of Sentiment Analysis through Social Media Platforms" discusses challenges in sentiment analysis of large, unorganized text data from social media, highlighting informal, noisy, and ambiguous content, and the impact of regional, cultural, and contextual language differences. "A Survey of Sentiment Analysis on Social Media" examines the difficulty in determining text relevance due to the conversational and noisy style of social media texts and explores language nuances, cultural differences, and contextual dependencies. "Sentiment Analysis on Social Media" reviews the comprehensive challenges and methodologies in sentiment analysis but lacks empirical validation or specific case studies. Future studies should explore new methods or address existing issues in sentiment analysis. Including real-life examples and data could enhance credibility and usefulness.

Results

Omuya and Okeyo employ PCA and SVD for dimensionality reduction, preprocessing social media text with NLP techniques before classifying tweets using Naive Bayes or Support Vector Machines. Rodriguez-Ibanez and Moreno review multiple sentiment analysis methods (SVMs and RNNs), and evaluate their performance using metrics such as accuracy, precision, recall, and F1-score, as detailed in Table 3.

Table-3: Comparison of different analysis methods by Omuya and Okeyo

Methodology	Accuracy	Precision	Recall	F1- Score
Lexicon based	0.70	0.68	0.72	0.70
Machine learning (SVM)	0.75	0.74	0.73	0.75
Deep learning (LSTM)	0.82	0.81	0.83	0.82
Hybrid Approach	0.85	0.84	0.86	0.85

Methodology section of Ying et al. (2020) could detail common sentiment analysis techniques on social media, including data gathering, feature extraction, and classification algorithms like Support Vector Machines (SVM) and Recurrent Neural Networks (RNN). They might also evaluate model effectiveness metrics and existing social sentiment analysis tools, addressing the complexity and dynamic nature of social media content. Their article covers lexicon-based strategies, machine learning algorithms, and hybrid approaches, emphasizing the significance of sentiment analysis in brand monitoring, political analysis, and customer feedback, as shown in Table-4.

Table-4: Comparison of analysis techniques by Ying et al. (2020)

Technique	Accuracy	Computational Cost	Data Requirement	Handling Sarcasm	Context Awareness	Scalability
SVM	High (80-90%)	Medium	High	Low	Medium	High
Naïve Bayes	Moderate (70-80%)	Low	Medium	Low	Low	High
Lexicon based	Variable (50-80%)	Low	Low	Low	Low	Medium
Deep learning	Very high (85-95%)	High	Very high	Medium	High	Medium
Hybrid Method	High (80-90%)	High	High	Medium	Medium	Medium

Review of Social Media Sentiment Analysis Methods

Yadav's (2023) research highlights practical sentiment analysis applications in brand monitoring, political orientation assessment, and customer feedback management. She underscores the maturity of social media environments and their real-life applications, noting the field's growth since 1994. Sentiment analysis operates on three levels: Sentence, Document, and Feature Level, analyzing extensive online information for better understanding and connection. It focuses on product ratings in e-commerce, encompassing emotions and attitudes through NLP and ML algorithms, with challenges and impacts on decision-making explored. Multimodal sentiment analysis helps gauge user opinions on various subjects, offering advantages in predicting box office success, election outcomes, and reading trends. Machine Learning (MLA) and Lexicon-Based Approaches (LBA) are compared, with MLA using specific algorithms and LBA examining positive and negative terms. Naive Bayes is essential for processing large datasets due to its efficient classification algorithms. Conditional probability provides insights into the likelihood of events under specific conditions.

$$p(X|Y) = \frac{p(Y|X)P(X)}{P(Y)}$$

(Eq.1)

The probabilities of any two events are calculated by multiplying their probabilities and dividing by their weighting factors. We randomly selected 100,000 Positive and 100,000 Negative tweets, calculating probabilities for Negation, Intensity, and combined functions. Sentiment Classification Based on Observation: A new lexicon-based sentiment analysis algorithm was evaluated for security, social media analytics, and longer messages like movie reviews, using the same word sets in both datasets. Five sentiment analysis algorithms were assessed:

1. L: Considers negation and boosting factors to encode sentiment scores for positive and negative words. It sums all sentiment values (positive/negative/neutral/zero) within sentences, classifying messages based on whether the value is positive, negative, or neutral.

2. LN: Similar to L but includes Eq. 2 steps, combining positive emotions with some negative ones. (Eq.2) IP and IN represent intensifiers for positive and negative terms, respectively. Instead of altering sentiment values by 50% or 100%, we modify word counts by increments of 0.5-1 per sentiment value increment. After integration and standardization, two values are obtained: one scale from 0-100 for positive sentiment and -100 to 0 for negative sentiment. The algorithm classifies tweets as positive or negative based on the greater absolute sentiment value.

$$F_P = min\{\frac{A_p}{2 - log(3.5 * W_P + I_P)}, 100\}$$

$$F_N = max\{\frac{A_p}{2 - log(3.5 * W_N + I_N)}, 100\} \qquad \text{(Eq.2)}$$

3. LNS: Analyzes each sentence and calculates sentiment by averaging all values.

4. LNW: Similar to LN but uses an evidence-based function (Eq) to categorize messages positively or negatively as shown below code:

```
If (wN==0)
     Return (finalsentiment (FP ,eP))
Else If (wP==0)
     Return (finalsentiment (FN ,eN))
Else {
     If (FP-FN> 0.1)
          Return (finalsentiment (FP ,eP))
     Else if (FN-FP> 0.1)
          Return (finalsentiment (FN ,eN))
     Else {
          If (FP+FN> 0)
                    Return (finalsentiment (FP ,eP))
          Else if (FN-FP< 0)
                    Return (final sentiment (FN ,eN))
          Else
                    Return (0)
          }
     }
int finalsentiment (F,e)
{
          if (|F| > 25) ||  |e| > 0.5)
               Return F
          else
               Return 0
}
```

5. LNWS: Applies LNW to each sentence and repeats the procedure from Figure 2. By averaging all sentence scores, the overall sentiment is determined. This article discusses the results of evaluating two datasets, gaining further insight by assessing the top-performing method on each dataset by precision, recall, and F-measure scores.

Review of Social Media Sentiment Analysis Methods

The paper examines current literature related to deep learning methods and multilingual sentiment analysis exhaustively. This methodical approach allows for in-depth investigation while maintaining reliability during the review process. The authors likely found significant studies, collected their main results, and organized them cohesively. To achieve this work's primary objectives, there are four steps

(i) Define the scope of the research
(ii) Searching for articles
(iii) Verification of articles
(iv) Analysis of research

Discussion

Iglesias and Moreno (2019) discussed lexicon-based approaches, machine-learning techniques, deep learning models, and hybrid approaches for sentiment analysis. Lexicon-based methods match words to infer sentiment, while machine learning techniques, such as Support Vector Machines (SVM), train on labeled data. Deep learning models like Convolution Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks learn complex data patterns. Hybrid approaches combine lexicon-based and machine learning elements. Their study "Sentiment Analysis for Social Media" highlights the importance of analyzing social media data to understand public opinion, market trends, and social dynamics, using a large dataset of social media posts from platforms like Twitter and Facebook.

Ha et al. (2019) utilize multilevel semantic network visualization to display emotional words in movie reviews through three methods: semantic word nodes, scaling maps of semantic data, and cluster visualizations using asterism images. These visualizations enable recommender systems to suggest movies with similar emotional content, potentially applicable across various social media platforms.

Kim and Jeong's paper (Kim & Jeong, 2019) "Sentiment Classification with Convolutional Neural Networks" evaluates their framework using three datasets: movie reviews, customer reviews, and the Stanford Sentiment Treebank. Their CNN model, with two convolutional layers, pooling layers, and fully connected layers, outperforms traditional Machine Learning and state-of-the-art Deep Learning models, highlighting the benefits of consecutive convolutional layers.

Mao et al. (2019) study "Using Sentiment-Aware Word Embedding for Emotion Classification" combines emotional word embeddings with Word2Vec semantic embeddings, forming a hybrid representation. Utilizing the DUTIR emotional lexicon, their method categorizes vocabulary into seven emotions. Evaluation on Weibo data demonstrates that hybrid word vectors effectively classify emotions under supervision. Jabreel and Moreno (2019) propose a multi-class emotion classification approach using deep learning. They introduce the xy-pair-set technique to convert multi-class problems into binary tasks, solvable by their BNet model. This system, featuring three embedding models with attention functions and RNN-based encoding modules, outperformed existing systems in SemEval-2019 Task 1 "Affect in Tweets."

Research Contribution

Iglesias and Moreno (2019) enhance sentiment analysis, improving real-time sentiment classification accuracy and reliability for these sources by adapting models trained on them. They contribute to the understanding of sentiment in social media, both in research and application. Omuya and Okeyo address emotion analysis challenges in social media tweets. "Sentiment analysis algorithms using dimensionality reduction and NLP techniques" were combined to improve accuracy and efficiency. Their methods and insights are invaluable for analysts, researchers interested in sentiment across platforms.

Rodriquez-Ibanez and Moreno's study (2023) on sentiment analysis in social media covers various practices, topologies, and research avenues in this field, providing a valuable reference for researchers and professionals. Liu (2012) extend this exploration by addressing key challenges and methods in sentiment analysis, suggesting future research directions. Their work is essential for those involved in social media research. The paper analyzes sentiment analysis on social media (2015-2022), examining dependencies and connections across platforms using Natural Language Processing algorithms to abstract sentiment from social media posts.

Applications of Sentiment Analysis in Social Media

Applications of sentiment analysis have grown, aiding the understanding of social events, product promotions, and political happenings. Four articles highlight its use in improving health insurance, studying AIDS patients, creating user profiles, and detecting cyber-attacks. The study "Using Social Media to Deduce Consumer Sentiment for Health Insurance during Enrollment Season" uses Twitter polling data and the NRC Emotion Lexicon to analyze consumer sentiment, revealing concerns about provider networks, prescription drug benefits, and politics. Healthcare professionals were trusted, but customers expressed anxiety about unexpected situations, suggesting insurers need more studies to tailor suitable policies.

Park and Woo (Park & Woo, 2019) focus on sentiment analysis of health-related discussions on internet forums, analyzing posts from HealthBoard.com for gender classification using traditional and deep learning methods. Liu et al.'s "Psychographic Segmentation Based on Online Review-Enhanced Recommendations" (Liu et al., 2019) explores whether psychographic traits can predict user preferences in e-commerce. They develop psychographic word lists using psycholinguistics and Word Net synonyms, creating good and bad word lists for the Schwartz Value Survey and Big Five Factor models. Word representations were generated using Word2Vec modeling and Amazon corpus word embeddings. They integrated word representations into a recommender system predicting online shopping habits using deep neural networks. Customers are segmented with BDSCAN, but this does not significantly improve results, and predictive power remains low.

Psychographic variables help explain e-consumer preferences. Gutierrez-Esparza et al.'s study "Classification of Cyber-Aggression Cases Using Machine Learning" (Gutiérrez-Esparza et al., 2019) identifies cyber aggression by classifying a large dataset of Latin American cyberbullying articles using machine learning, providing a corpus for non-English studies.

Table-5: Overview of sentiment analysis approaches in social media

Approach	**Description**	**Achievements**	**Challenges**	**Techniques**
Machine Learning	Uses Supervised learning algorithms	High accuracy with large datasets	Require Labeled data computationally intensive	SVM, Deep Learning, Naïve Bayes
Lexicon Based	Relies on a predefined set of words(lexicon)	No need for labeled data, easy to implement	Limited by the lexicon and struggle with context and sarcasm	SentiWordNet, AFINN, VADER
Hybrid Approaches	Combination of ML and Lexicon based techniques	Balance accuracy and generalization	Implementation Complexity, Coordination Problem	Hybrid Sentiment Analysis
Deep Learning	Uses CNN, RNN and LSTM for sentiment analysis	Handle complex language structure with high performance	Require large datasets, issue in overfitting	CNN, RNN, LSTM
Topic based sentiment analysis	Analyzes sentiments w.r.t. specific topics / aspects of the text.	Useful for multi-faceted data	Complex preprocessing, limited to topic modeling	LDA- based Sentiment Analysis

New Addition in sentiment analysis

In "Sentiment Analysis for Social Media," the focus is on selecting appropriate techniques and developing effective evaluation criteria. As part of promoting a healthy learning and living environment, free Wi-Fi access is recommended. The paper "Sentiment Analysis on Tweets from Social Media using Dimensionality Reduction and Natural Language Processing" outlines a sentiment analysis method using NLP, dimensionality reduction, and machine learning. It combines principal component analysis and independent component analysis for dimensionality reduction and highlight essential features, preprocessing data and analyzing sentiments using NLP. The model expands parts of speech to include verbs, adjectives, and adverbs, resulting in labeled tokens facilitating sentiment analysis. Evaluated using Naive Bayes (NB), support vector machines (SVM), and K-nearest neighbors, this model outperforms two other sentiment analysis models. The proposed model achieves two objectives: reducing feature dimensions for training and improving sentiment analysis accuracy by tagging relevant labeled tokens with parts of speech. An analysis of deep learning and multilingual sentiment on social media provides insight into their intersection. A comprehensive analysis of methodologies and challenges is needed, adding value to existing literature and resources for researchers and practitioners interested in sentiment analysis.

Conclusion

The objective of this review is to investigate the collection and analysis methods used in various research papers related to sentiment analysis / opinion mining, enabling us infer from all the found resources a comprehensive state-of-the-art measurement for expression-based learning. Our results show that sentiment analysis has applications in several practical scenarios. To summarize, sentiment analysis is an impactful technique and can reveal numerous insights for various applications. Sentiment analysis is an emerging interdisciplinary field at the crossroads of both computer science and psychology, so let me put it this way sentiment analysis is human behavior tracker. So, sentiment analysis is to be examined in detail so that human behavior can be analyzed and understood better. In this way, this review paper is able to provide a complete snapshot of the latest and future development on sentiment analysis opinion mining.

Acknowledgements

All the data is taken from reliable sources and is properly mentioned with references.

Conflict of Interest: There is no affiliation or involvement in any funding organization.

Funding statement: No funding source

Ethical consideration: The study is approved by the ethics committee of Qurtba University.

Author Contributions: The Muhammad Tufail[2] contributed in determining methodology and Taj Rehman[3] contributed in data simulation and analysis.

References

Abulwafa, A. (2022). A Survey of Deep Learning Algorithms and its Applications. *Nile Journal of Communication and Computer Science*, *0*(0), 0–0. https://doi.org/10.21608/njccs.2022.139054.1000

Alotaibi, F. M. (2023). A Machine-Learning-Inspired Opinion Extraction Mechanism for Classifying Customer Reviews on Social Media. *Applied Sciences (Switzerland)*, *13*(12). https://doi.org/10.3390/app13127266

Amangeldi, D., Usmanova, A., & Shamoi, P. (2024). Understanding Environmental Posts: Sentiment and Emotion Analysis of Social Media Data. *IEEE Access*, *12*. https://doi.org/10.1109/ACCESS.2024.3371585.

Amangeldi, D., Usmanova, A., Shamoi, P., Go, A., Bhayani, R., Huang, L., Zhang, L., Wang, S., Liu, B., Fan, F., Feng, Y., Zhao, D., He, B., Zhou, D., Xiao, J., Jiang, X., Liu, Q., Yuan, N. J., Xu, T., … Potts, C. (2020). BERT-MK: Integrating graph contextualized knowledge into pre-trained language models. *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, *4*, 3667–3677. https://doi.org/10.18653/v1/2020.acl-main.338

Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 41, Issue 2). https://doi.org/10.1109/TPAMI.2018.2798607

Barik, K., & Misra, S. (2024). Analysis of customer reviews with an improved VADER lexicon classifier. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-023-00861-x

Başarslan, M. S., & Kayaalp, F. (2024). Sentiment analysis using a deep ensemble learning model. *Multimedia Tools and Applications*, *83*(14). https://doi.org/10.1007/s11042-023-17278-6

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *2020-December*.

Chen, X. W., & Lin, X. (2014). Big data deep learning: Challenges and perspectives. In *IEEE Access* (Vol. 2). https://doi.org/10.1109/ACCESS.2014.2325029

Cirqueira, D., Almeida, F., Cakir, G., Jacob, A., Lobato, F., Bezbradica, M., & Helfert, M. (2020). Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail. *CHIRA 2020 - Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications*. https://doi.org/10.5220/0010215303190328

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*.

Dos Santos, C. N., & Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. *COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers*.

Fan, F., Feng, Y., & Zhao, D. (2018). Multi-grained attention network for aspect-level sentiment classification. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 3433–3442. https://doi.org/10.18653/v1/d18-1380

Garg, S., Panwar, D. S., Gupta, A., & Katarya, R. (2020). A literature review on sentiment analysis techniques involving social media platforms. *PDGC 2020 - 2020 6th International Conference on Parallel, Distributed and Grid Computing*, *November*, 254–259. https://doi.org/10.1109/PDGC50313.2020.9315735

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12), 2009.

Gutiérrez-Esparza, G. O., Vallejo-Allende, M., & Hernández-Torruco, J. (2019). Classification of cyber-aggression cases applying machine learning. *Applied Sciences (Switzerland)*, *9*(9). https://doi.org/10.3390/app9091828

Halawani, H. T., Mashraqi, A. M., Badr, S. K., & Alkhalaf, S. (2023). Automated sentiment analysis in social media using Harris Hawks optimisation and deep learning techniques. *Alexandria Engineering Journal*, *80*. https://doi.org/10.1016/j.aej.2023.08.062

Hota, H. S., Sharma, D. K., & Verma, N. (2023). Integration of deep learning techniques for sentiment and emotion analysis of social media data. *International Journal of Intelligent Systems Technologies and Applications*, *21*(1). https://doi.org/10.1504/IJISTA.2023.130552

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*. https://doi.org/10.18653/v1/p18-1031

Ibañez, M. M., Rosa, R. R., & Guimarães, L. N. F. (2020). Sentiment analysis applied to analyze society’s emotion in two different context of social media data. *Inteligencia Artificial*, *23*(66). https://doi.org/10.4114/INTARTIF.VOL23ISS66PP66-84

Iglesias, C. A., & Moreno, A. (2019). Sentiment analysis for social media. In *Applied Sciences (Switzerland)* (Vol. 9, Issue 23). https://doi.org/10.3390/app9235037

Ismet, H. T., Mustaqim, T., & Purwitasari, D. (2022). Aspect Based Sentiment Analysis of Product Review Using Memory Network. *Scientific Journal of Informatics*, *9*(1). https://doi.org/10.15294/sji.v9i1.34094

Ha, H., Han, H., Mun, S., Bae, S., Lee, J., & Lee, K. (2019). An Improved Study of Multilevel Semantic Network Visualization for Analyzing Sentiment Word of Movie Review Data. *Applied Sciences*. https://doi.org/10.3390/app9122419.

Jabreel, M., & Moreno, A. (2019). A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets. *Applied Sciences*. https://doi.org/10.3390/app9061123.

Jain, P. K., Quamer, W., Saravanan, V., & Pamula, R. (2023). Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis. *Journal of Ambient Intelligence and Humanized Computing*, *14*(8). https://doi.org/10.1007/s12652-022-03698-z

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, *1*. https://doi.org/10.3115/v1/p14-1062

Khan, J., Ahmad, N., Khalid, S., Ali, F., & Lee, Y. (2023). Sentiment and Context-Aware Hybrid DNN With Attention for Text Sentiment Classification. *IEEE Access*, *11*. https://doi.org/10.1109/ACCESS.2023.3259107

Khan, L., Amjad, A., Afaq, K. M., & Chang, H. T. (2022). Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media. *Applied Sciences (Switzerland)*, *12*(5). https://doi.org/10.3390/app12052694

Kim, H., & Jeong, Y. S. (2019). Sentiment classification using Convolutional Neural Networks. *Applied Sciences (Switzerland)*, *9*(11). https://doi.org/10.3390/app9112347

Kim, Y. (2014). Convolutional neural networks for sentence classification. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. https://doi.org/10.3115/v1/d14-1181

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

Liu, H., Huang, Y., Wang, Z., Liu, K., Hu, X., & Wang, W. (2019). Personality or value: A comparative study of psychographic segmentation based on an online review enhanced recommender system. *Applied Sciences (Switzerland)*, *9*(10). https://doi.org/10.3390/app9101992

Mao, X., Chang, S., Shi, J., Li, F., & Shi, R. (2019). Sentiment-Aware Word Embedding for Emotion Classification. *Applied Sciences*. https://doi.org/10.3390/app9071334.

Miyato, T., Dai, A. M., & Goodfellow, I. (2017). Adversarial training methods for semi-supervised text classification. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.

Omuya, E. O., Okeyo, G., & Kimwele, M. (2023). Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. *Engineering Reports*, *5*(3). https://doi.org/10.1002/eng2.12579

Park, S., & Woo, J. (2019). Gender classification using sentiment analysis and deep learning in a health web forum. *Applied Sciences (Switzerland)*, *9*(6). https://doi.org/10.3390/app9061249

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*. https://doi.org/10.18653/v1/n18-1202

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. In *Expert Systems with Applications* (Vol. 223). https://doi.org/10.1016/j.eswa.2023.119862

Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE*, *16*(2), 1–23. https://doi.org/10.1371/journal.pone.0245909

Sánchez-Núñez, P., de las Heras-Pedrosa, C., & Peláez, J. I. (2020). Opinion mining and sentiment analysis in marketing communications: A science mapping analysis in web of science (1998-2018). *Social Sciences*, *9*(3). https://doi.org/10.3390/socsci9030023

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1631–1642.

Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics*, *14*(2). https://doi.org/10.4018/IJHISI.2019040101

T.k., B., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. In *Computer Science Review* (Vol. 40). Elsevier Ireland Ltd. https://doi.org/10.1016/j.cosrev.2021.100395

Tang, D., Qin, B., & Liu, T. (2016). Aspect level sentiment classification with deep memory network. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*. https://doi.org/10.18653/v1/d16-1021

van den Broek-Altenburg, E. M., & Atherly, A. J. (2019). Using social media to identify consumers' sentiments towards attributes of health insurance during enrollment season. *Applied Sciences (Switzerland)*, *9*(10). https://doi.org/10.3390/app9102035

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-December*.

Wu, C. H., & Li, T. Y. (2016). Social sensor: An analysis tool for social media. *International Journal of Electronic Commerce Studies*, *7*(1), 77–94. https://doi.org/10.7903/ijecs.1411

Wu, X., & Lode, M. (2020). Language Models are Unsupervised Multitask Learners (Summarization). *OpenAI Blog*, *1*(May).

Yadav, J. (2023). Sentiment Analysis on Social Media. *Qeios*, 1–10. https://doi.org/10.32388/yf9x04

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, *32*.

Ying, T., Jin, Z., & Lixin, X. (2020). A survey of sentiment analysis on social media. *Data Analysis and Knowledge Discovery*, *4*(1). https://doi.org/10.11925/infotech.2096-3467.2019.0769

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, *60*(2). https://doi.org/10.1007/s10115-018-1236-4

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

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2023). A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. *IEEE Transactions on Knowledge and Data Engineering*, *35*(11). https://doi.org/10.1109/TKDE.2022.3230975