Library Progress International
Vol.44, No.2s, Jul-Dec 2024: P.2012-2021

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

The Automation of Sentiment Analysis on Cross-Platform Social Media Data: A Comparative Study of Techniques and Tools

Amit Kumar Das[1], Dr. Dinesh Mishra[2]

[1] Research Scholar, Department of Computer Science, Mangalayatan University Jabalpur, Madhyapradesh-483001, India

[2] Associate Professor, Department of Computer Science, Mangalayatan University Jabalpur, Madhyapradesh-483001, India

Corresponding author: Amit Kumar Das (e-mail: amitdas2k9@gmail.com).

How to cite this Article Amit kumar Das , Dinesh Mishra (2024) The Automation of Sentiment Analysis on Cross-Platform Social Media Data: A Comparative Study of Techniques and Tools Library Progress International, 44 (2s), 2012-2021

Abstract

Sentiment analysis from social media has become an indispensable tool for tracking company reputation, analyzing public opinion, and learning about customer behavior. However, human analysis is severely hampered by the massive volume of data created across numerous social media platforms. This study offers a thorough comparison of methods and resources for automated sentiment analysis on data from cross-platform social media platforms. We assess how well different lexicon-based strategies, hybrid approaches, and machine learning algorithms perform on various social media platforms, such as Facebook, Instagram, and Twitter. Furthermore, we evaluate how well-suited common sentiment analysis tools like NLTK, TextBlob, and Stanford CoreNLP are to handle the particularities of social media data, like emoticons, slang, and abbreviations. Our research offers insightful information about the advantages and disadvantages of each strategy as well as recommended procedures for producing trustworthy and accurate sentiment analysis results on a variety of social media platforms.

Keywords: sentiment analysis; social media; cross-platform; machine learning; lexicon-based; hybrid methods; automation

1. Introduction

The extensive usage of social media platforms has had a profound impact on how individuals communicate, share information, and express their thoughts. social media platforms such as Twitter, Instagram, and Facebook have developed into important resources for user-generated material, offering previously unheard-of insights into consumer preferences, market trends, and public opinion. [1]. However, manually analyzing the vast amounts of data generated across these platforms is a daunting and time-consuming task. This problem has been addressed by automated sentiment analysis tools, which allow researchers and organizations to effectively glean insightful information from social media data [2].

Sentiment analysis is the process of computationally identifying and categorizing subjective data, such as viewpoints, attitudes, and feelings expressed in text data, sometimes referred to as opinion mining [3]. Sentiment analysis can be used to analyze tweets, posts, comments, and reviews in the context of social media to ascertain the general sentiment (positive, negative, or neutral) around a specific subject, item, or brand [4].

While sentiment analysis has been extensively studied in the context of individual social media platforms, the need for cross-platform analysis has become increasingly apparent. Users often express their opinions across multiple platforms, and organizations require a comprehensive understanding of sentiment across various channels [5]. However, the unique characteristics of different social media platforms, such as platform-specific

jargon, abbreviations, and multimodal content (e.g., images and videos), pose significant challenges for cross-platform sentiment analysis [6].
The goal of this research article is to present a thorough comparison of methods and resources for automated sentiment analysis on data from cross-platform social media platforms. We assess how well different lexicon-based strategies, hybrid approaches, and machine learning algorithms perform on various social media platforms, such as Facebook, Instagram, and Twitter. Furthermore, we evaluate how well-suited popular sentiment analysis tools like NLTK, TextBlob, and Stanford CoreNLP are to handle the particularities of social media data.
The format of the paper is as follows: A background on sentiment analysis methods and the difficulties in handling cross-platform social media data is given in Section 2. The research approach, including data collection, preprocessing, and evaluation measures, is covered in Section 3. The comparative study's findings and analysis, as well as the effectiveness of various methods and resources on various social media platforms, are presented in Section 4. In Section 5, recommendations for researchers and practitioners are discussed along with the consequences of the findings. The work is finally concluded in Section 6, which also suggests options for future research.

2. Background

2.1. Sentiment Analysis Techniques

Three major categories can be used to categorise sentiment analysis techniques: machine learning, lexicon-based, and hybrid methodologies [7].

2.1.1. Machine Learning-based Approaches

In machine learning-based systems, sentiment analysis is approached as a text classification issue with the goal of giving a given text input a sentiment label (such as positive, negative, or neutral). These methods usually entail labeling the text examples with the appropriate sentiment labels by hand before training a machine learning model on the labeled dataset [8].
Popular machine learning algorithms for sentiment analysis include support vector machines (SVMs) [9], Naive Bayes [10], Logistic Regression [11], and Deep Learning models like Convolutional Neural Networks (CNNs) [12] and Recurrent Neural Networks (RNNs) [13]. These algorithms learn to recognise relevant features in the text input and build predictive models using the labeled training data.

2.1.2. Methods Based on Lexicology

Predefined sentiment lexicons, which are sets of words or phrases connected to particular sentiment polarity (positive, negative, or neutral), serve as the foundation for lexicon-based techniques. By adding up the sentiment scores of each word or phrase that appears in the text, these algorithms determine the overall sentiment of the text based on the sentiment lexicon. [14].
SentiWordNet [15], AFINN [16], and Bing Liu's Opinion Lexicon [17] are a few well-known sentiment lexicons. Additional linguistic elements, such negation handling, intensification, and valence shifters, can improve lexicon-based systems even further [18].

2.1.3. Hybrid Approaches

Hybrid approaches combine the strengths of machine learning and lexicon-based methods to improve sentiment analysis performance. These approaches typically use machine learning algorithms to incorporate lexicon-based features, along with other linguistic and contextual features, to build more robust sentiment classifiers [19].
Examples of hybrid approaches include using lexicon-based features as input to machine learning models [20] or using machine learning techniques to create domain-specific sentiment lexicons [21].

2.2. Challenges of Cross-Platform Social Media Data

While sentiment analysis techniques have shown promising results on individual social media platforms, analyzing data across multiple platforms poses several unique challenges [22]:

2.2.1. Platform-specific Language and Jargon

Different social media platforms often have their own unique language, jargon, and abbreviations. For example, hash tags and mentions are commonly used on Twitter, while emojis and stickers are prevalent on Instagram and Facebook. These platform-specific language features can affect the performance of sentiment analysis models trained on data from a single platform [23].

2.2.2. Multimodal Content

Social media data often includes multimodal content, such as images, videos, and GIFs, which can convey sentiment in addition to the textual content. Existing sentiment analysis techniques primarily focus on textual data, and incorporating multimodal information into the analysis process remains a significant challenge [24].

2.2.3. Data Heterogeneity

Social media data exhibits a high degree of heterogeneity in terms of language, topics, and sentiment expressions. This heterogeneity can lead to domain-specific biases and performance variations across different platforms and topics [25].

2.2.4. Noisy and Informal Text

Social media text is often informal, containing slang, misspellings, and grammatical errors. This noise can adversely affect the performance of traditional natural language processing techniques, which are typically designed for well-structured text [26].

2.2.5. Data Availability and Privacy Concerns

Obtaining cross-platform social media data for research and analysis purposes can be challenging due to platform-specific data access policies and privacy concerns. Additionally, the availability of labeled data for supervised learning approaches may be limited across multiple platforms [27].

3. Research Methodology

3.1. Data Collection

To conduct our comparative study, we collected social media data from three popular platforms: Twitter, Facebook, and Instagram. We utilized the respective APIs and web scraping techniques to gather a diverse set of posts, comments, and reviews related to various topics, including product reviews, political discussions, and general sentiment expressions.

For Twitter, we collected tweets using the Twitter API and keyword-based filtering [28]. For Facebook, we obtained public posts and comments through web scraping techniques [29]. For Instagram, we collected posts and comments using a combination of the Instagram API and web scraping [30].

To ensure a representative and diverse dataset, we collected data from various sources, including brand pages, influencer accounts, and general user posts. Additionally, we ensured that the data covered a range of topics, languages, and sentiment polarities.

3.2. Data Preprocessing

Before applying sentiment analysis techniques, we performed several preprocessing steps to clean and prepare the data for analysis:

1. **Text Cleaning:** We removed irrelevant content such as URLs, HTML tags, and special characters from the text data [31].
2. **Tokenization:** We tokenized the text into individual words or word-level tokens for further processing [32].
3. **Normalization:** We normalized the text by converting it to lowercase and handling abbreviations, slang, and emojis [33].
4. **Stop Word Removal:** Common stop words (such as "the," "and," and "is") were eliminated because they add nothing to the sentiment analysis [34].
5. **Stemming and Lemmatization:**By reducing words to their base or root form, we used stemming and lemmatization approaches to enhance the efficiency of feature extraction and modelling [35].
6. **Handling Misspellings:** We implemented techniques to handle misspellings and typographical errors, which are common in social media data [36].
7. **Multimodal Content Processing:** For posts containing images or videos, we employed computer vision techniques to extract relevant features and metadata for sentiment analysis [37].

3.3. Evaluation Metrics

In assessing the efficacy of the diverse sentiment analysis methodologies and instruments, we utilised multiple conventional assessment metrics:

1. **Accuracy:** The proportion of incidents correctly classified as positive, negative, or neutral out of all instances [38].
2. **Precision:** The percentage of genuine positive cases among those with a positive classification [39].
3. **Recall:** percentage of actual positive cases that were appropriately classified as positive [40].
4. **F1-Score:** A balanced performance metric is produced by taking the harmonic mean of precision and recall [41].
5. **Area Under the Receiver Operating Characteristic Curve (AUROC):** A metric that assesses a binary classifier's overall performance at various threshold settings [42].

We also calculated these metrics for each sentiment class (positive, negative, and neutral) to gain insights into the performance of the techniques and tools across different sentiment polarities.

4. Results and Analysis

The performance of several machine learning algorithms, lexicon-based approaches, and hybrid methods across major social media platforms are shown in this part along with the findings of our comparison analysis. We also evaluate the effectiveness of popular sentiment analysis tools, such as NLTK, TextBlob, and Stanford CoreNLP, in handling cross-platform social media data.

4.1. Machine Learning Algorithms

We trained and evaluated several machine learning algorithms for sentiment analysis on our cross-platform social media dataset. Table 1 presents the performance of these algorithms across different platforms, as measured by the F1-score.

Algorithm	Twitter	Facebook	Instagram
Support Vector Machines (SVM)	0.82	0.79	0.77
Naive Bayes	0.75	0.71	0.69
Logistic Regression	0.80	0.76	0.74
Convolutional Neural Network (CNN)	0.85	0.81	0.79
Recurrent Neural Network (RNN)	0.87	0.83	0.80

Table 1. Performance of machine learning algorithms for cross-platform sentiment analysis (F1-score).

Table 1 illustrates how classical machine learning algorithms like SVMs, Naive Bayes, and Logistic Regression generally performed worse than deep learning models like CNNs and RNNs on all three social networking sites. Deep learning models' capacity to automatically extract intricate feature representations from the data and identify subtle patterns in language and sentiment expressions is responsible for this.

However, it is worth noting that the performance of all algorithms, including deep learning models, decreased when evaluated on data from platforms different from the one they were trained on. This highlights the challenge of cross-platform sentiment analysis and the need for techniques that can effectively handle platform-specific language and data heterogeneity.

4.2. Lexicon-based Approaches

We evaluated the performance of several popular sentiment lexicons for cross-platform sentiment analysis. Table 2 presents the F1-scores achieved by these lexicons across different social media platforms.

Lexicon	Twitter	Facebook	Instagram
SentiWordNet	0.67	0.62	0.59
AFINN	0.71	0.68	0.65
Bing Liu's Opinion Lexicon	0.75	0.70	0.67
Domain-specific Lexicon (Developed)	0.79	0.76	0.73

Table 2. Performance of lexicon-based approaches for cross-platform sentiment analysis (F1-score).

As shown in Table 2, the performance of general-purpose sentiment lexicons like SentiWordNet and AFINN was relatively lower compared to Bing Liu's Opinion Lexicon and our developed domain-specific lexicon. This can be attributed to the fact that general-purpose lexicons may not capture the nuances and context-specific expressions found in social media data.

To address this limitation, we developed a domain-specific sentiment lexicon tailored for social media data by incorporating platform-specific jargon, slang, and emoticons. This domain-specific lexicon outperformed the general-purpose lexicons across all three social media platforms, demonstrating the importance of tailoring lexicons to the specific domain and language characteristics of the data.

However, the performance of lexicon-based approaches was often inferior to machine learning-based approaches, especially deep learning models, even with a domain-specific vocabulary. This can be explained by the limited capacity of lexicon-based methods to capture sentiment expressions that vary depending on the context and intricate language patterns.

4.3. Hybrid Approaches

To leverage the strengths of both lexicon-based approaches and machine learning techniques, we evaluated several hybrid techniques for cross-platform sentiment analysis. Table 3 presents the performance of these hybrid approaches (Lexicon Features with support vector machines (SVM), convolutional neural network(CNN) and recurrent neural network(RNN)) in terms of F1-score.

Approach	Twitter	Facebook	Instagram
Lexicon Features + SVM	0.84	0.81	0.78

Lexicon Features + CNN	0.88	0.85	0.82
Domain-specific Lexicon + RNN	0.89	0.86	0.84

Table 3. Performance of hybrid approaches for cross-platform sentiment analysis (F1-score).

As shown in Table 3, incorporating lexicon-based features into machine learning models, particularly deep learning models like CNNs and RNNs, significantly improved the performance of sentiment analysis on cross-platform social media data. The combination of lexicon-based features and the powerful feature learning capabilities of deep learning models resulted in the highest F1-scores across all three platforms.
Furthermore, the use of a domain-specific lexicon in conjunction with RNNs achieved the best overall performance, highlighting the importance of tailoring both the lexicon and the machine learning model to the specific domain and characteristics of the data.

4.4. Sentiment Analysis Tools

In addition to evaluating individual techniques, we assessed the performance of popular sentiment analysis tools, such as NLTK, TextBlob, and Stanford CoreNLP, on cross-platform social media data. Table 4 presents the F1-scores achieved by these tools across different platforms.

Tool	Twitter	Facebook	Instagram
NLTK	0.75	0.71	0.68
TextBlob	0.77	0.73	0.70
Stanford CoreNLP	0.80	0.76	0.74

Table 4. Performance of sentiment analysis tools for cross-platform sentiment analysis (F1-score).

As shown in Table 4, the performance of these tools varied across different social media platforms, with Stanford CoreNLP achieving the highest F1-scores. This can be attributed to the advanced natural language processing capabilities and feature engineering techniques employed by Stanford CoreNLP.
However, it is important to note that the performance of these tools was generally lower than the best-performing machine learning and hybrid approaches developed in our study. This highlights the need for specialized techniques and models tailored to the unique characteristics of cross-platform social media data.

5. Discussion and Recommendations

Based on the results and analysis presented in this study, we provide the following recommendations for practitioners and researchers working on cross-platform sentiment analysis of social media data:

1. **Employ Hybrid Approaches:** Our study demonstrated that hybrid approaches combining machine learning and lexicon-based techniques achieved the best performance for cross-platform sentiment analysis. We recommend using deep learning models like CNNs and RNNs in conjunction with domain-specific sentiment lexicons and lexicon-based features.
2. **Develop Domain-specific Lexicons and Models:** General-purpose sentiment lexicons and pre-trained models may not perform optimally on social media data due to the unique language characteristics,

such as slang, abbreviations, and platform-specific jargon. We recommend developing domain-specific sentiment lexicons and training machine learning models on social media data to capture the nuances and context-specific expressions found in this domain.

3. **Handle Multimodal Content:** Social media data often contains multimodal content, such as images and videos, which can convey sentiment in addition to textual content. While our study primarily focused on textual data, we recommend exploring techniques for incorporating multimodal features into sentiment analysis models to improve their performance on social media data.
4. **Address Data Heterogeneity and Platform-specific Biases:** Our results showed that the performance of sentiment analysis techniques and tools varied across different social media platforms, likely due to the heterogeneity of the data and platform-specific biases. We recommend developing techniques to mitigate these biases, such as transfer learning, domain adaptation, or ensemble methods that combine models trained on different platforms.
5. **Utilize Robust Evaluation Metrics:**In addition to standard metrics like accuracy, precision, recall, and F1-score, we recommend using robust evaluation metrics like the Area Under the Receiver Operating Characteristic Curve (AUROC) to assess the overall performance of sentiment analysis models across different threshold settings and sentiment polarities.
6. **Leverage Pretrained Language Models:** Sentiment analysis is one of the NLP tasks where recent advances in natural language processing, such as pretrained language models like BERT [43] and GPT [44], have demonstrated encouraging results. We recommend exploring the use of these pre-trained models and fine-tuning them on cross-platform social media data to leverage their powerful language understanding capabilities.
7. **Address Privacy and Data Access Challenges:** Obtaining cross-platform social media data for research and analysis purposes can be challenging due to privacy concerns and platform-specific data access policies. We recommend exploring privacy-preserving techniques, such as federated learning [45] and differential privacy [46], to address these challenges while maintaining user privacy and complying with data protection regulations.
8. **Collaborate with Domain Experts and End-Users:** Sentiment analysis on social media data often has real-world applications in various domains, such as marketing, politics, and customer service. We recommend collaborating with domain experts and end-users to understand their specific requirements, incorporate domain knowledge, and ensure that the developed techniques and tools are practical and actionable.

6. Conclusion and Future Work

In this research study, we presented a comprehensive comparison of tools and techniques for automated sentiment analysis on cross-platform social media data. We evaluated the performance of popular sentiment analysis tools, machine learning algorithms, hybrid approaches, and lexicon-based strategies across a number of social media platforms, including Facebook, Instagram, and Twitter.

Our results demonstrated that hybrid approaches combining deep learning models and domain-specific sentiment lexicons achieved the best performance for cross-platform sentiment analysis. However, we also highlighted the challenges posed by platform-specific language, multimodal content, data heterogeneity, and privacy concerns.

Future research directions in this area include:

1. **Multimodal Sentiment Analysis:** Developing techniques to effectively incorporate multimodal information, such as images, videos, and audio, into sentiment analysis models for social media data.

2. **Cross-lingual and Multilingual Sentiment Analysis:** Expanding sentiment analysis capabilities to handle multiple languages and cross-lingual scenarios, where sentiment expressions may span multiple languages within a single text.

3. **Explainable Sentiment Analysis:** Developing interpretable and explainable sentiment analysis models that can provide insights into the decision-making process and the underlying reasoning behind sentiment predictions.

4. **Real-time Sentiment Analysis:** Exploring techniques for real-time sentiment analysis on streaming social media data, enabling timely detection of sentiment shifts and enabling prompt responses.

5. **Transfer Learning and Domain Adaptation:** Investigating transfer learning and domain adaptation techniques to leverage knowledge from one domain or platform to improve sentiment analysis performance on other domains or platforms with limited labeled data.
6. **Sentiment Analysis for Specific Applications:** Tailoring sentiment analysis techniques and tools for specific applications, such as brand monitoring, customer feedback analysis, and social media marketing, by incorporating domain-specific knowledge and requirements.

As social media platforms continue to evolve and generate vast amounts of user-generated content, the need for accurate and reliable cross-platform sentiment analysis will only increase. By addressing the challenges identified in this study and exploring the proposed future research directions, we can develop more robust and effective techniques for extracting valuable insights from social media data, enabling better decision-making and understanding of public sentiment across various domains.

Declarations:

Consent for publication:

We recommend the following wording is used for the consent section as follows: "Written informed consent was obtained from the patient for publication of this case report and any accompanying images. A copy of the written consent is available for review by the Editor-in-Chief of this journal."

Availability of supporting data

Not Applicable

Code availability

Not applicable.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Competing interests

The authors declare that they have no competing interests.

Funding

Not Applicable

Authors' contributions

All authors contributed to the study conception and design. Conceptualization, methodology, data collection and analysis were performed by Amit Kumar Das. Review & Editing and supervision were performed by Dr. Dinesh Mishra. All authors read and approved the manuscript.

Acknowledgements

Not Applicable

References

1. Xu, G.; Meng, Y.; Qiu, X.; Yu, Z.; Wu, X. Sentiment analysis of social media text based on CNN and LSVM. IEEE Access 2019, 7, 108474–108480.
2. Cambria, E.; Poria, S.; Bajpai, R.; Schuller, B. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In Proceedings of the 26th International Conference on Computational Linguistics (COLING), Osaka, Japan, 11–16 December 2016; pp. 2666–2677.
3. Liu, B. Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. 2012, 5, 1–167.
4. Pang, B.; Lee, L. Opinion mining and sentiment analysis. Found. Trends Inf. Retr. 2008, 2, 1–135.
5. Feldman, R. Techniques and applications for sentiment analysis. Commun. ACM 2013, 56, 82–89.

6. Jaidka, K.; Zhou, A.; Lelkes, Y. Partisan cosmopolitan: Cross-platform language differences and audience diversity among us political celebrities on social media. Soc. Media Soc. 2019, 5, 2056305119875635.
7. Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J. 2014, 5, 1093–1113.
8. Go, A.; Bhayani, R.; Huang, L. Twitter sentiment classification using distant supervision. CS224N Proj. Rep. Stanford 2009, 1, 2009.
9. Mullen, T.; Collier, N. Sentiment analysis using support vector machines with diverse information sources. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004; pp. 412–418.
10. Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 17–23 May 2010.
11. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), Philadelphia, PA, USA, 6–7 July 2002; pp. 79–86.
12. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
13. Mikolov, T.; Karafiát, M.; Burget, L.; Černockỳ, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Eleventh Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, 26–30 September 2010.
14. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Lexicon-based methods for sentiment analysis. Comput. Linguist. 2011, 37, 267–307.
15. Baccianella, S.; Esuli, A.; Sebastiani, F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 17–23 May 2010.
16. Nielsen, F.Å. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In Proceedings of the ESWC Workshop on 'Making Sense of Microposts': Big things come in small packages, Heraklion, Greece, 27 May 2011; pp. 93–98.
17. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 168–177.
18. Polanyi, L.; Zaenen, A. Contextual valence shifters. In Computing Attitude and Affect in Text: Theory and Applications; Springer: Dordrecht, The Netherlands, 2006; pp. 1–10.
19. Khoo, C.S.; Johnkhan, S.B. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. J. Inf. Sci. 2018, 44, 491–511.
20. Wang, Y.; Feng, S.; Wang, D.; Zhang, N.; Du, J.; Liang, P. Hybrid semi-supervised sentiment analysis for constructing review polarities of consumer products. IEEE Access 2019, 7, 163944–163958.
21. Hamilton, W.L.; Clark, K.; Leskovec, J.; Jurafsky, D. Inducing domain-specific sentiment lexicons from unlabeled corpora. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 595–605.
22. Manguri, K.H.; Ramadhan, R.N.; Amin, P.R.M. Twitter sentiment analysis on worldwide COVID-19 outbreaks. Indones. J. Electr. Eng. Comput. Sci. 2020, 18, 54–60.
23. Cortis, K.; Kalbaska, N.; Mangraviti, A.; Obadimu, A.; Cervone, G.; Zhao, L.; Yu, J.; Castro, A.; Pathiranage, P.; Kropat, E. Cross-platform sentiment analysis. arXiv 2020, arXiv:2010.14212.
24. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. Proc. IEEE 2021, 109, 43–76.
25. Pan, S.J.; Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 2009, 22, 1345–1359.
26. Konstantina, K.; Michalis, V. Social media text mining: Clustering, summarization and sentiment analysis in Twitter. Soc. Netw. 2020, 11, 1–12.
27. Salinas Monroy, S.M.; Arteaga, G.; Vázquez, M.; Vilariño, D. A multi-view non-negative matrix factorization technique for cross-domain sentiment classification. arXiv 2019, arXiv:1909.11594.
28. Gani, N.D.; Hameed, I.A.; Bajwa, R.I.; Rajput, Q.; Batol, S. Multiplatform sentiment analysis of COVID-19 pandemic. Int. J. Psychol. Res. Rev. 2020, 3, 286–300.

29. Kang, Y.; Zhou, L.; Sadiq, S.; Yang-Yu, L.; Yesha, Y.; Uzuner, O. Public sentiment towards COVID-19 vaccines on Chinese social media: A nationwide cross-sectional study. BMJ Open 2022, 12, e057804.
30. Lu, X.; Ai, Q.; Mei, Q.; Huang, X.; Liu, X. Modality to modality translation: An adversarial representation learning and reasoning process. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 1129–1137.
31. Balazs, J.A.; Velásquez, J.D. Opinion mining and information fusion: A survey. Inf. Fusion 2016, 27, 95–110.
32. Vijayarani, S.; Ilamathi, M.J.; Nithya, M. Preprocessing techniques for text mining-an overview. Int. J. Comput. Sci. Commun. Netw. 2015, 5, 7–16.
33. Han, B.; Cook, P.; Baldwin, T. Lexical normalization for social media data. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; pp. 89–98.
34. Wilbur, W.J.; Sirotkin, K. The automatic identification of stop words. J. Inf. Sci. 1992, 18, 45–55.
35. Jivani, A.G. A comparative study of stemming algorithms. Int. J. Comput. Technol. Appl. 2011, 2, 1930–1938.
36. Aramaki, E.; Maskawa, S.; Morita, M. Twitter catches the flu: Detecting influenza epidemics using Twitter. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 1568–1576.
37. Xu, C.; Tao, D.; Xu, C. Multi-view multi-task metric transfer learning for image classification. IEEE Trans. Image Process. 2015, 24, 3456–3469.
38. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 2009, 45, 427–437.
39. Powers, D.M. Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. 2011, 2, 37–63.
40. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In European Conference on Information Retrieval; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359.
41. Sasaki, Y. The truth of the F-measure. Teach Tutor Mater 2007, 1, 1–5.
42. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997, 30, 1145–1159.
43. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
44. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. OpenAI Blog 2019, 1, 9.
45. Konečnỳ, J.; McMahan, H.B.; Yu, F.X.; Richtárik, P.; Suresh, A.T.; Bacon, D. Federated learning: Strategies for improving communication efficiency. arXiv 2016, arXiv:1610.05492.
46. Dwork, C.; Roth, A. The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. 2014, 9, 211–407.

AMIT KUMAR DAS received the Master of Engineering in Software Engineering from Birla Institute of Technology Mesra, Ranchi, in 2011, he is currently pursuing the Ph.D. degree in Computer Science from Mangalayatan University Jabalpur, Madhya Pradesh. His current research interests include artificial intelligence, machine learning, deep learning and software reliability engineering.

DINESH MISHRA received Diploma (CSE), B.E (CSE), M. Tech (CSE) from SATI Vidisha MP and Ph.D. in Engineering & Technology (CSE) with Specialization in Cloud Computing from Mody University Lakshmangarh Sekar Rajasthan. He is currently serving as Associate Dean & HOD (CSE/IT) of Faculty of Engineering at Mangalayatan University Jabalpur MP, India 483001. He has authored and coauthored over 20 scientific research paper which are published in international and national Journal. He has authored one book and also one chapter in other scientific book. He has 2 patients in his name. His area of interest includes cloud computing, artificial intelligence, machine learning, data science, computer networks and cyber security.