Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Harnessing Deep Learning to Combat Misinformation and Detect Depression on Social Media: Challenges and Interventions

Swati Lokhande, Dr. Daya Shankar Pandey

Department of Computer Application, RKDF Ist, SRK University, India
E-mail: swatilokhande29@gmail.com

In today's digital landscape, social media platforms have revolutionized human interaction, fundamentally altering how individuals and entities connect, communicate, and disseminate information. This paper investigates the multifaceted repercussions of social media on contemporary society, examining its influence on cultural norms, political discourse, commercial endeavors, and activist movements. Despite its capacity to foster connectivity and communal engagement, social media presents inherent challenges, notably the proliferation of harmful misinformation, exemplified by the surge in vaccine hesitancy. Navigating these challenges necessitates a nuanced approach that reconciles the principles of free expression with the imperative of safeguarding public health. Moreover, the research methodology section introduces a deep learning framework for identifying depression signals within social media content, offering a pathway to developing effective interventions to mitigate adverse effects.

Keywords: Social media, Impact, Misinformation, Vaccine hesitancy, Depression detection, Deep learning.

1. Introduction

Social media has revolutionized the way individuals and organizations connect, communicate, and share information in the digital age. With the advent of platforms like facebook, Twitter, Instagram, LinkedIn, and others, people across the globe are seamlessly interconnected, transcending geographical barriers to form virtual communities and networks. Social media platforms serve as dynamic spaces for users to express themselves, exchange ideas, and engage in a myriad of activities ranging from personal interactions to professional networking, content creation, and marketing. This digital ecosystem has not only reshaped how we talk but has also influenced various factors of society, including culture, politics, commerce, and activism. As social media continues to evolve and integrate into daily life, its impact and significance in

shaping human interactions and societal dynamics are simple, making it a subject of massive interest and scrutiny in present day discourse.

Social media, while providing an unprecedented capacity for the public to communicate, has also been a major factor in the rise of fringe opinions damaging to public health [1]. Reconciling principles of free speech with the policing of social media for damaging falsehoods remains a conundrum for democracies. Vaccine hesitancy is not a new phenomenon, but the proliferation of anti-vaccination misinformation through social media has given it new urgency, especially in light of the coronavirus pandemic and hopes for rapid development and deployment of a vaccine [3].

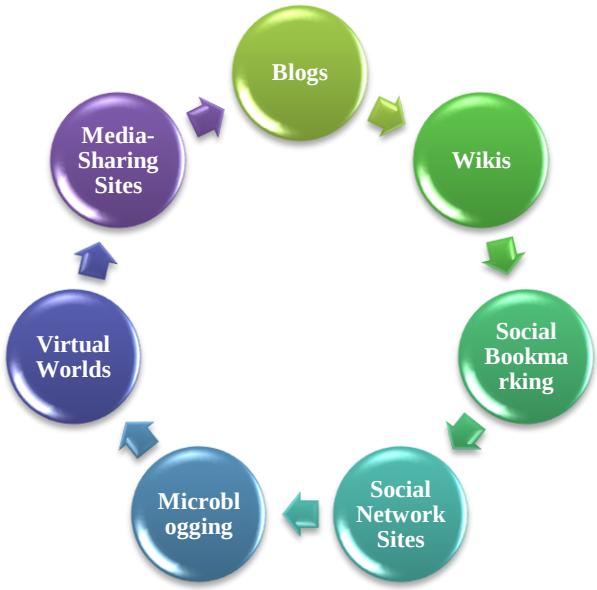

Fig. 1 Types of social media

Fig. 1 shown in Social media encompasses various online platforms that facilitate communication, content creation, and interaction. From blogs like WordPress to collaborative wikis such as Wikipedia, these platforms enable individuals to share thoughts, expertise, and information. Social network sites like Facebook and Twitter connect users within virtual communities, while microblogging platforms like Twitter allow for brief updates and engagement. Virtual worlds like Second Life offer immersive experiences, and media-sharing sites like YouTube and Instagram enable the sharing and discovery of multimedia content [3]. Together, these platforms shape the digital landscape, fostering connectivity, creativity, and community online.

The Impact of Social Media

What impact does social media have on modern society? Because their widespread use did not begin until the early 2000s, both the economic and social effects of these channels of communication are still not completely understood. Although several believe that the World Wide Web is further alienating people, others believe it will increase democratic participation.

Based on analysts, we are unlikely to encounter either a euphoric society with content social media users or an unfavourable society of loners. On the contrary, we are confronted with a less unified civilization than we are accustomed to. Additionally, there do not appear to be any discernible differences in the total amount of social ties or the level of social involvement amongst Internet users with non-users. On the contrary, the Internet, particularly social media, is providing avenues for encouraging involvement in the community as well as strengthening bonds. The use of social media has had a financial impact on materials creation as well as utilisation, as well as the IT and telecommunications business. Furthermore, numerous companies are incorporating social media into their marketing strategies. Social media provide more channels for governments and parliamentarians to communicate with the public [4]. For example, the Health Agency of Canada employs a variety of social media platforms to distribute information about public health issues. A growing number of legislators across the nation are making use of social networking sites.

2. Literature Review

Tong et al. [5] investigated information influence maximization in competitive environments, utilizing the HITS algorithm for data preprocessing in social networks. Through experiments on real-world Twitter datasets, they proposed an algorithm showcasing improved accuracy and computation time compared to existing methods. Future research directions include incorporating temporal considerations and enhancing algorithm efficiency for topic-based influence maximization. Pran et al. [6] analyzed Bangladeshi individuals' sentiment towards coronavirus-related news on Facebook, employing deep learning algorithms to classify emotions into three categories: Analytical, Depressed, and Angry. Their findings reveal a predominant analytical sentiment among commenters, providing insights into public psychology amidst the pandemic. Malarvizhi [7] explored sentiment categorization in Social Networking Sites (SNSs), evaluating sentiment analysis (SA) and document mining techniques (DMTs) for accuracy and execution speed. Their study demonstrates DMTs' effectiveness in categorizing sentiments, achieving an accuracy of 92%. Alquran and Banitaan [8] investigated features for fake news detection, constructing prediction models using Naive Bayes, Bayesian Network, and J48 classification methods. Their experiments on a benchmark dataset yielded an overall F-score of 69.7%, with the J48 classifier performing best on politicians' statements. Marengo et al. [9] demonstrated the predictability of addictive tendencies towards social media based on digital footprints of Facebook users using machine learning strategies. Their ensemble models achieved sufficient accuracy rates in distinguishing between disordered and non-disordered social media users, suggesting the inference of individual differences in social media addiction from digital traces. Sundararaj et al. [10] conducted a customer review analysis to understand changing behaviors towards various products. Their study revealed that product quality varied depending on customer profiles, with factors such as high quality, social media influence, and customer profiles significantly impacting purchasing behavior.

Table 1 Performance Assessment of Algorithms and Models in Social Media Research

Study	Algorithm	Parameter(s)	Values	Result
Deng et al. [11]	Frequent pattern mining-based algorithm	Precision, Recall Rate, F1 Measure	Precision: 90.9%, F1: 85.8%, Recall: 76.5%	Proposed algorithm enhances precision, recall rate, and F1 measure.
Subramani et al. [12]	GBDT-CGBO method	GBDT-precision, Recall, F-score, RMSE, Accuracy	GBDT-precision: 96.92%, Recall: 90.45%, F-score: 90.12%, RMSE: 32.91%, Accuracy: 94.65%	GBDT-CGBO method proves effective for Influencers Prediction.
Belcastro et al. [13]	HASHET model	Correct Classification Percentage, Incorrect and Neutral Classifications	Correct classification: Up to 77%	HASHET model excels in uncovering main hashtag-based discussion topics.
Pourhabibi et al. [14]	Unsupervised Laplacian Score based approach	Accuracy	Over 88%	Proposed approach demonstrates high accuracy in spammer detection.
Nistor et al. [15]	Advanced machine learning methods	Accuracy	Fake news detection accuracy: 98.88%	Advanced ML methods improve detection of fake news.
Li et al. [16]	Supervised machine learning based solution	F1 Score	F1 values: 89.79%, 86.78%, 86.24% on three ground truth datasets	UGC-based user identification model achieves high performance.
Jin et al. [17]	Data mining technology	Information Search Behavior	Improved information search efficiency and effectiveness	Data mining technology enhances understanding of information search behavior.
Ho et al. [19]	Various classification algorithms	Performance Analysis, Classification Algorithms	Random forest: 98%, Decision tree: 90%, AdaBoost: 89%, LR: 88%, SVM: 86%, SGD: 84%	Random forest achieves highest score in educational data mining analysis.

Table 1 shown in various novel approaches and algorithms have been proposed and tested across diverse domains, each showcasing remarkable improvements in specific performance metrics. Deng et al. [11] introduce an algorithm that enhances precision, recall rate, and F1 measure, with notable achievements such as precision reaching 90.9%. Subramani et al. [12] demonstrate the effectiveness of the GBDT-CGBO method for Influencers Prediction, achieving impressive metrics such as GBDT-precision of 96.92% and an accuracy of 94.65%. Belcastro et al. [13] find that the HASHET model excels in uncovering main hashtag-based discussion topics, with correct classification percentage reaching up to 77%. Pourhabibi et al.'s [14] approach for spammer detection surpasses 88% accuracy. Nistor et al.'s [15] advanced ML methods significantly enhance fake news detection accuracy, achieving an impressive accuracy rate of 98.88%. Li et al.' [16] UGC-based user identification model achieves high performance in F1 score, ranging from 86.24% to 89.79% across three ground truth datasets. Jin et al. [17] demonstrate improved information search efficiency and effectiveness through data mining technology. Kumar et al.' [18] transformers-based RoBERTa model outperforms existing approaches in mental stress identification. Ho et al. [19] performance analysis of classification algorithms in educational data mining analysis shows random forest achieving the highest score of 98%, followed by decision tree (90%), AdaBoost (89%), LR (88%), SVM

(86%), and SGD (84%).

3. Research Methodology

Classifying depression from social media content can be a challenging task, but it can be done using various techniques and tools. One approach is to use a multimodal approach, which involves analyzing multiple modalities of social media content, such as text, images, and videos. This chapter is dedicated to design a methodology for multi-modular and multi-lingual depression classification using deep learning approach. In this section, the methodology is designed in four steps as presented in Fig. 2.

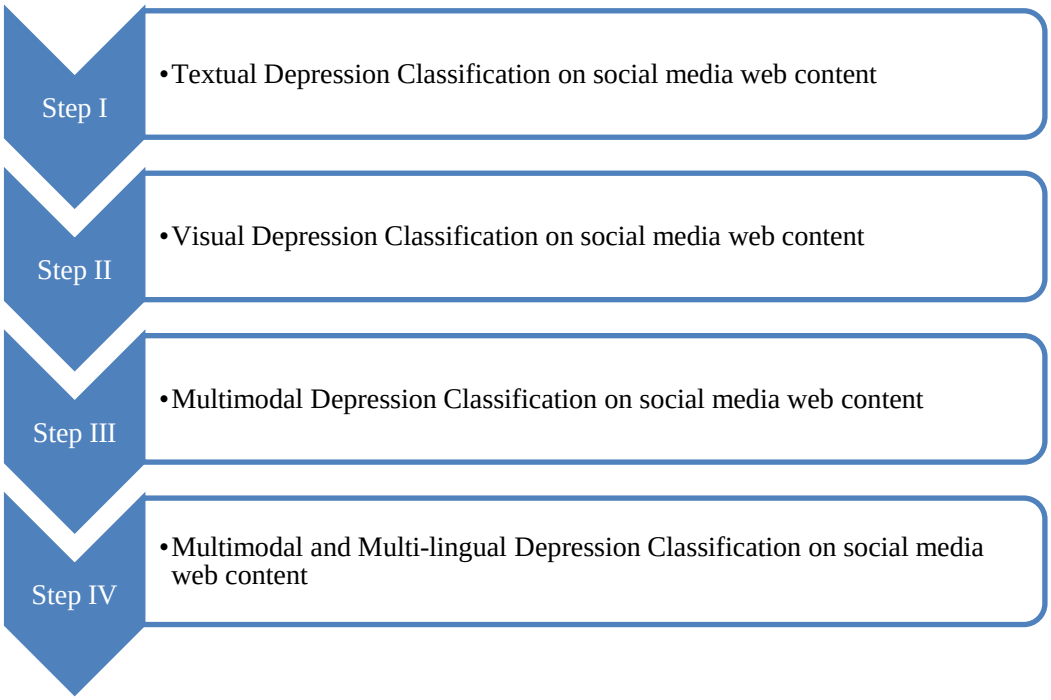

Fig. 2: Steps of Methodology

Textual Depression Classification on social media web content

In this section, textual depression classification is descibed using a deep learning model that has three main components. The components are outlined in Fig. 3.

- The first component is constructing a word vector. This involves converting the text data into a numerical vector representation that can be used by the model.
- The second component is using a BERT (Bidirectional Encoder Representations from Transformers) model. BERT is a type of deep learning model that is designed to understand the context of words in a sentence. The model will use this understanding to identify aspects related to depression and extract features at the aspect level.

- The third component is passing the extracted features into a classifier to determine whether the text is indicative of depression. The classifier is a machine learning model that will make a decision based on the extracted features.

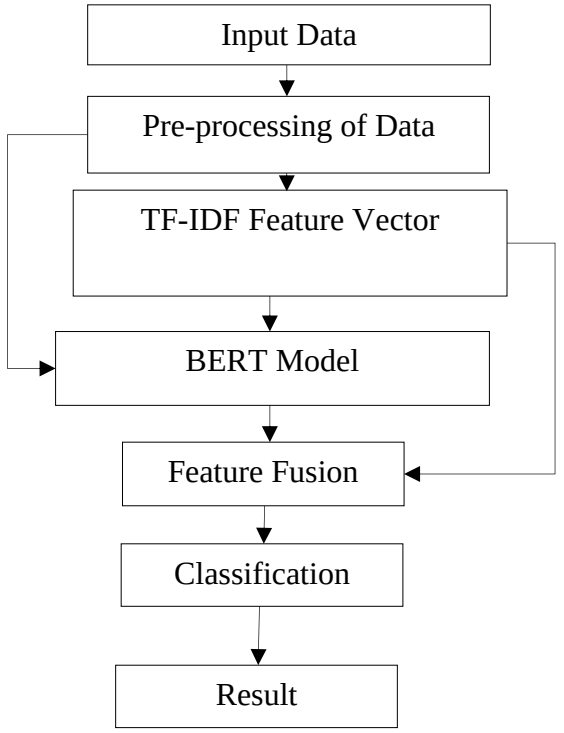

Fig. 3: Proposed Flow Diagram for Textual Depression Classification

4. Result and Discussion

Result Analysis of Textual Social Media Content Depression Classification

The table 2 presents the results for five classifiers: Random Forest, Linear SVM, KNeighbors, Gradient Boosting, and Decision Tree. From the table, we can see that the GradientBoosting classifier achieved the highest accuracy score of 94.00%, followed by Random Forest and KNeighbors with accuracy scores of 92.00% and 91.00%, respectively. The highest precision score of 93.40% was achieved by the Random Forest classifier, while the highest recall score of 92.00% was achieved by the GradientBoosting classifier. The F1-scores show a similar pattern to precision and recall, with the GradientBoosting classifier achieving the highest score of 92.50%, followed by RandomForest, DecisionTree, KNeighbors, and LinearSVM. Overall, the results suggest that Gradient Boosting and Random Forest classifiers perform well on the textual content using TF-IDF features.

Table 2: Performance Analysis of Textual Content on TF-IDF Features

Classifiers	Accuracy	Precision	Recall	F1-score
Random Forest	92.00%	93.40%	90.00%	91.67%
Linear SVM	89.00%	91.00%	88.00%	89.47%
KNeighbors	91.00%	91.00%	91.00%	91.00%
GradientBoosting	94.00%	93.00%	92.00%	92.50%
DecisionTree	91.50%	92.00%	92.00%	92.00%

Figure 5.1 shows the accuracy analysis of textual content on TF-IDF features. The figure shows that the Gradient Boosting classifier achieved the highest accuracy score of 94%, followed by Random Forest and Decision Tree with accuracy scores of 92% and 91.5%, respectively. LinearSVM and k-NN achieved the lowest accuracy scores of 89% and 91%, respectively.

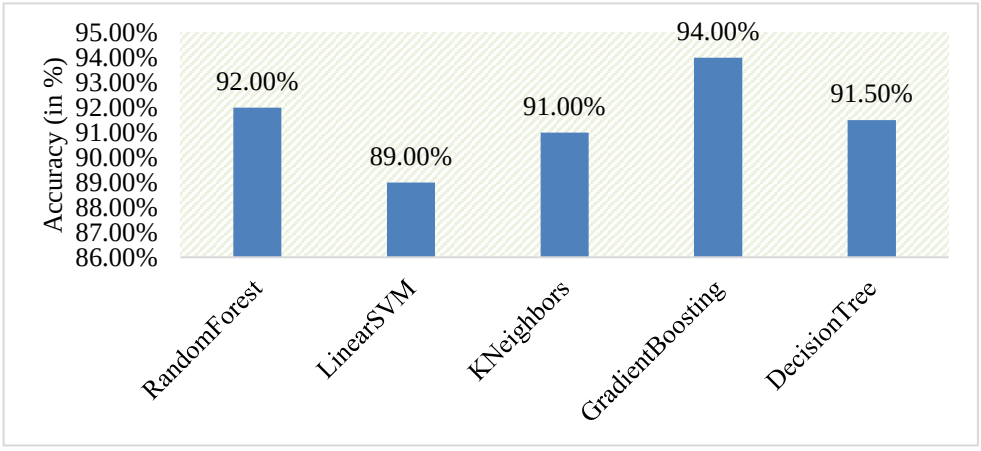

Fig. 4: Accuracy Analysis of Textual Content on TF-IDF Features

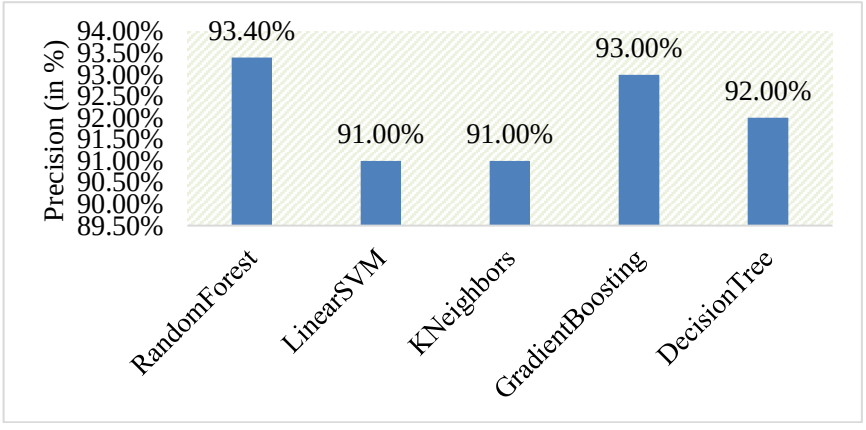

Fig. 5: Precision Analysis of Textual Content on TF-IDF Features

Figure 5.2 shows the precision analysis of textual content on TF-IDF features. The figure shows that the random forest classifier achieved the highest precision score of 93.4%, followed by Gradient boositng and Decision Tree with precision scores of 93% and 92%, respectively. LinearSVM and k-NN achieved the lowest precision scores of 91% and 91%, respectively.

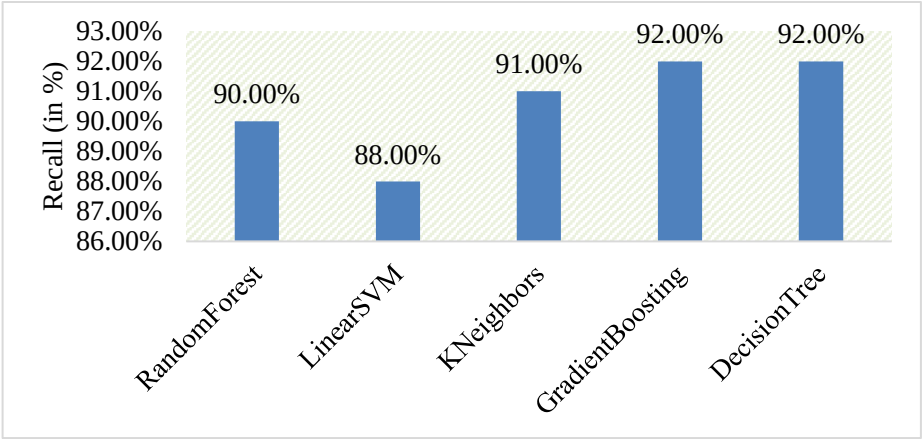

Fig. 6: Recall Analysis of Textual Content on TF-IDF Features

Figure 5.3 shows the Recall analysis of textual content on TF-IDF features. The figure shows that the random forest classifier, LinearSVM, KNeighbors , Gradient boosting and decision tree achieved the Recall scores of 90%, 88%, 91%, 92%, 92 %, respectively, in which highest is 92% of decision tree.

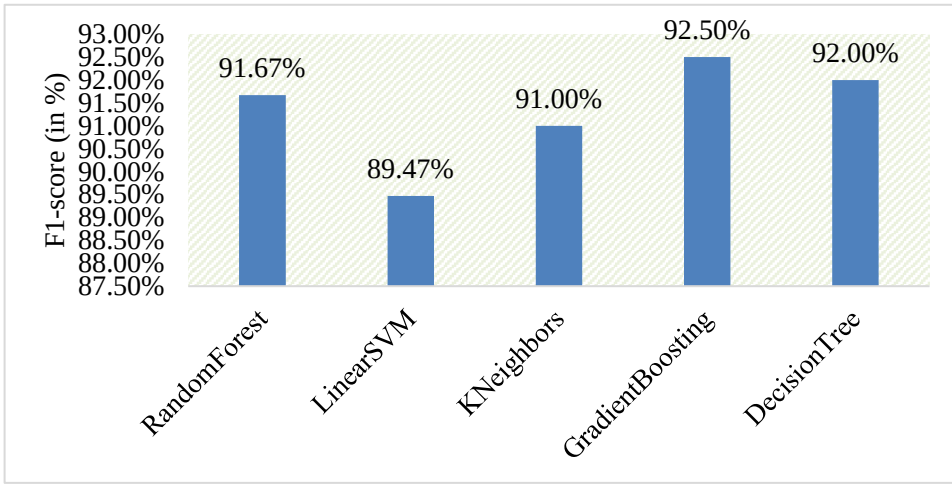

Fig. 7: F1_score Analysis of Textual Content on TF-IDF Features

Figure 5.4 shows the F-1 Score analysis of textual content on TF-IDF features. The figure shows that the random forest classifier, LinearSVM, KNeighbors , Gradient boosting and decision tree achieved the F-1 Score scores of 91.66%, 89.4%, 91%, 92.5% and 90%, respectively, in which highest is 92.5% of decision tree.

5. Conclusion

Social media stands as a pivotal force in contemporary society, reshaping communication paradigms and redefining social dynamics. Yet, alongside its transformative potential, it

presents formidable obstacles, including the dissemination of false information and its detrimental impact on public health, as evidenced by the rise of vaccine hesitancy. Addressing these challenges requires a delicate balance between upholding the principles of free expression and curtailing the spread of harmful content. The development of innovative methodologies, such as deep learning algorithms for depression detection from social media data, represents a crucial step toward understanding and mitigating the negative ramifications of social media. Going forward, interdisciplinary collaboration and creative solutions are imperative to harnessing the constructive power of social media while safeguarding individual and collective well-being.

References

1. Wilson, S. L., & Wiysonge, C. (2020). Social media and vaccine hesitancy. BMJ global health, 5(10), e004206.
2. Belcastro, L.; Cantini, R.; Marozzo, F. Knowledge Discovery from Large Amounts of Social Media Data. Appl. Sci. 2022, 12, 1209. https://doi.org/10.3390/app12031209
3. A. Kumar, T. E. Trueman and E. Cambria, "Stress Identification in Online Social Networks," 2022 IEEE International Conference on Data Mining Workshops (ICDMW), Orlando, FL, USA, 2022, pp. 427-434, doi: 10.1109/ICDMW58026.2022.00063.
4. Yongjun Li, Zhen Zhang, You Peng, Hongzhi Yin, Quanqing Xu, Matching user accounts based on user generated content across social networks, Future Generation Computer Systems, Volume 83, 2018, Pages 104-115, ISSN 0167-739X, https://doi.org/10.1016/j.future.2018.01.041.
5. J. Tong, L. Shi, L. Liu, J. Panneerselvam and Z. Han, "A novel influence maximization algorithm for a competitive environment based on social media data analytics," in Big Data Mining and Analytics, vol. 5, no. 2, pp. 130-139, June 2022, doi: 10.26599/BDMA.2021.9020024.
6. M. S. A. Pran, M. R. Bhuiyan, S. A. Hossain and S. Abujar, "Analysis Of Bangladeshi People's Emotion During Covid-19 In Social Media Using Deep Learning," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225500.
7. Malarvizhi, A. "An assessment of data mining techniques reliability in predicting social media sentiments." Journal of Positive School Psychology (2022): 7254-7263.
8. H. Alquran and S. Banitaan, "Fake News Detection in Social Networks Using Data Mining Techniques," 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2022, pp. 155-160, doi: 10.1109/AIIoT54504.2022.9817287.
9. Marengo D, Montag C, Mignogna A, Settanni M. Mining Digital Traces of Facebook Activity for the Prediction of Individual Differences in Tendencies Toward Social Networks Use Disorder: A Machine Learning Approach. Front Psychol. 2022 Mar 8;13:830120. doi: 10.3389/fpsyg.2022.830120. PMID: 35350734; PMCID: PMC8957912.
10. Vinu Sundararaj, M R Rejeesh, A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites, Journal of Retailing and Consumer Services, Volume 58, 2021, 102190, ISSN 0969-6989, https://doi.org/10.1016/j.jretconser.2020.102190
11. K. Deng, L. Xing, L. Zheng, H. Wu, P. Xie and F. Gao, "A User Identification Algorithm

Based on User Behavior Analysis in Social Networks," in IEEE Access, vol. 7, pp. 47114-47123, 2019, doi: 10.1109/ACCESS.2019.2909089.

12. Subramani, N.; Veerappampalayam Easwaramoorthy, S.; Mohan, P.; Subramanian, M.; Sambath, V. A Gradient Boosted Decision Tree-Based Influencer Prediction in Social Network Analysis. Big Data Cogn. Comput. 2023, 7, 6. https://doi.org/10.3390/bdcc7010006
13. Belcastro, L.; Cantini, R.; Marozzo, F. Knowledge Discovery from Large Amounts of Social Media Data. Appl. Sci. 2022, 12, 1209. https://doi.org/10.3390/app12031209
14. Pourhabibi, T., Boo, Y.L., Ong, KL., Kam, B., Zhang, X. (2019). Behavioral Analysis of Users for Spammer Detection in a Multiplex Social Network. In: , et al. Data Mining. AusDM 2018. Communications in Computer and Information Science, vol 996. Springer, Singapore. https://doi.org/10.1007/978-981-13-6661-1_18
15. Nistor, A.; Zadobrischi, E. The Influence of Fake News on Social Media: Analysis and Verification of Web Content during the COVID-19 Pandemic by Advanced Machine Learning Methods and Natural Language Processing. Sustainability 2022, 14, 10466. https://doi.org/10.3390/su141710466
16. Yongjun Li, Zhen Zhang, You Peng, Hongzhi Yin, Quanqing Xu, Matching user accounts based on user generated content across social networks, Future Generation Computer Systems, Volume 83, 2018, Pages 104-115, ISSN 0167-739X, https://doi.org/10.1016/j.future.2018.01.041.
17. Jin, H., Miao, Y., Jung, JR. et al. Construction of information search behavior based on data mining. Pers Ubiquit Comput 26, 233–245 (2022). https://doi.org/10.1007/s00779-019-01239-8
18. A. Kumar, T. E. Trueman and E. Cambria, "Stress Identification in Online Social Networks," 2022 IEEE International Conference on Data Mining Workshops (ICDMW), Orlando, FL, USA, 2022, pp. 427-434, doi: 10.1109/ICDMW58026.2022.00063.
19. Ho, Chaang-Iuan, Ming-Chih Chen, and Ya-Wei Shih. "Customer engagement behaviours in a social media context revisited: using both the formative measurement model and text mining techniques." Journal of Marketing Management 38.7-8 (2022): 740-770.