JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Comparative Study of Machine Learning Techniques for Hate Speech Detection on Social Media Platforms

Indu Bala[1], Dr. Ikvinderpal Singh[2]

[1]Research Scholar, Department of Computer Science & Applications, Arni University, Kathgarh, Indora(H.P.)
[2]Assistant Professor, PG Department of Computer Science & Applications, Trai Shatabdi GGS Khalsa College, Sri Amritsar Sahib (Punjab).

ABSTRACT

Hate speech on social media platforms poses significant challenges in maintaining a safe and inclusive online environment. Automated hate speech detection using machine learning techniques has emerged as a promising solution. This paper presents a comparative study of three popular machine learning algorithms: Support Vector Machines (SVM), Random Forest, and Logistic Regression, for hate speech detection. Each algorithm is implemented and trained using the preprocessed data and hyper parameter tuning is performed to optimize their performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to measure the effectiveness of the models. The comparative study's contributions lie in its performance evaluation, methodological guidance, practical implementation insights, dataset considerations, and insights for model selection. Overall, this comparative study advances the understanding of hate speech detection techniques and provides guidance for selecting appropriate machine learning algorithms in real-world applications.

Keywords

Machine learning; Fake news; Support vector machine, Random forest; Logistic regression; Social media;

1. INTRODUCTION

Social media platforms have become integral parts of our daily lives, connecting millions of people worldwide. They provide platforms for communication, information sharing, and community engagement. However, alongside the benefits of social media, there is a growing concern regarding the proliferation of hate speech on these platforms. Hate speech refers to any form of communication, whether written, spoken, or symbolic, that promotes violence, discrimination, or hostility towards individuals or groups based on attributes such as race, ethnicity, religion, gender, sexual orientation, or other protected characteristics.

The problem of hate speech on social media platforms is a pressing issue that has gained significant attention in recent years. The rise of online hate speech has had far-reaching consequences, impacting individuals, communities, and societies as a whole. Hate speech not only undermines the fundamental principles of equality, diversity, and inclusion but also contributes to the perpetuation of stereotypes, discrimination, and social divisions.

One of the key challenges in addressing hate speech on social media platforms is the sheer volume and speed at which content is generated. Millions of messages, comments, and posts are shared every day, making it virtually impossible for human moderators to manually review and moderate every instance of hate speech. Consequently, there is a critical need for automated methods to detect and combat hate speech effectively.

Automated hate speech detection systems leverage machine learning techniques to identify and classify potentially offensive or harmful content on social media platforms. These systems aim to provide an early warning system, enabling timely intervention to mitigate the negative impact of hate speech. By automatically flagging and filtering hate speech content, these systems can help create safer and more inclusive online environments.

However, detecting hate speech on social media platforms is a complex and challenging task. Hate speech can take various forms, ranging from explicit slurs and derogatory language to subtle and coded expressions. Additionally, hate speech often involves nuanced contextual understanding, sarcasm, and cultural references, making it difficult to rely solely on keyword-based approaches. Furthermore, the dynamic nature of language and the evolving strategies employed by hate speech perpetrators require adaptive and robust detection systems.

To address these challenges, researchers and practitioners have turned to machine learning techniques, such as SVM, random forest, and logistic regression, to develop automated hate speech detection models. These models leverage large datasets of labeled hate speech and non-hate speech content to learn patterns and features that can distinguish between the two. Comparative studies of different machine learning techniques play a crucial role in understanding the strengths and limitations of these approaches and guiding the development of more effective hate speech detection systems.

2. OBJECTIVE OF THE STUDY

The objective of this study is to conduct a comparative analysis of machine learning techniques for the detection of hate speech on social media platforms. The study aims to evaluate the performance of three specific techniques: Support Vector Machines (SVM), Random Forest, and Logistic Regression. By comparing these techniques, the study seeks to identify their strengths and weaknesses in detecting hate speech, providing insights into their suitability and effectiveness for this specific task. By comparing these three machine learning techniques, the study aims to provide insights into their respective performance, strengths, and limitations for hate speech detection. The results of the study will help researchers, practitioners, and platform administrators make informed decisions regarding the selection and implementation of appropriate machine learning techniques for combating hate speech on social media platforms.

3. LITERATURE REVIEW

Existing research on hate speech detection using machine learning techniques has witnessed significant advancements in recent years. Researchers have explored various approaches and models to tackle the complex problem of identifying and mitigating hate speech on social media platforms. Here is an overview of the key research areas and findings in this domain:

1. *Feature-Based Approaches:* Early research focused on defining and extracting relevant features from text data to train hate speech detection models. These features include lexical, syntactic, and semantic information, as well as contextual and stylistic cues. Techniques such as n-grams, bag-of-words, and sentiment analysis were employed to capture hate speech characteristics. Researchers found that these features, when combined with traditional machine learning algorithms, achieved reasonable accuracy in hate speech detection.
2. *Deep Learning Models:* With the rise of deep learning, researchers explored the use of neural network architectures for hate speech detection. Recurrent Neural Networks (RNNs), specifically variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), were employed to capture sequential dependencies in text data. Convolutional Neural Networks (CNNs) were also utilized to model local patterns and hierarchical representations. These deep learning models demonstrated improved performance compared to traditional methods, especially in capturing nuanced hate speech patterns.
3. *Transfer Learning and Pretrained Language Models:* Pretrained language models, such as BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pretrained Transformer), and RoBERTa, have revolutionized hate speech detection. These models are trained on massive amounts of text data and can be fine-tuned on hate speech detection tasks. Transfer learning from these models has shown significant improvements in hate speech detection accuracy, as they capture rich contextual information and semantic relationships.
4. *Ensemble Learning:* Ensemble learning techniques, including Random Forest, Gradient Boosting, and Voting Classifier, have been employed to combine multiple models for hate speech detection. By leveraging the diversity of individual models, ensemble methods can enhance overall performance and mitigate the biases of individual classifiers. Researchers found that ensemble approaches often achieve higher accuracy and robustness in hate speech detection tasks.
5. *Multilingual Hate Speech Detection:* Hate speech detection is not limited to a single language, as it occurs across various cultural and linguistic contexts. Researchers have explored approaches for multilingual hate speech detection, leveraging cross-lingual transfer learning, language-agnostic features, and parallel corpora. These studies aim to address the challenges of detecting hate speech in languages with limited labeled data and diverse linguistic characteristics.
6. *Adversarial Attacks and Countermeasures:* As hate speech detection models are deployed, adversaries can attempt to circumvent them by generating adversarial examples. Researchers have investigated adversarial attacks on hate speech detection models and proposed countermeasures, including adversarial training, input perturbation, and robust model architectures. This line of research aims to enhance the resilience of hate speech detection models against malicious manipulation.
7. *Ethical Considerations and Bias Mitigation:* Hate speech detection models need to be developed with ethical considerations in mind, ensuring fairness and mitigating biases. Researchers have examined bias in hate speech datasets and developed methods to reduce biases during training and testing stages. This research emphasizes the importance of creating models that do not perpetuate or amplify existing biases in hate speech detection.

Existing research on hate speech detection using machine learning techniques has made significant progress in improving accuracy, scalability, and robustness. The use of deep learning models, pretrained language models, ensemble learning, and ethical considerations has advanced the field. However, challenges such as context sensitivity, cultural variations, and bias mitigation continue to be areas of active research, paving the way for more effective and responsible hate speech detection systems.

4. METHODOLOGY

A. Dataset:

The dataset used in a comparative study for hate speech detection can greatly influence the outcomes and generalizability of the findings. It's important to note that the availability and quality of annotated datasets can vary, and different studies may use different datasets based on their specific research objectives. Furthermore, some studies may involve domain-specific or multilingual datasets to cater to specific research needs or to address hate speech detection challenges in different contexts.
When conducting a comparative study, researchers ensure that the datasets used for different machine learning techniques are representative of the problem domain and exhibit a balanced distribution of hate speech and non-hate speech instances.

B. Preprocessing Steps:

Preprocessing steps play a crucial role in preparing text data for hate speech detection. These steps involve transforming raw text into a format suitable for machine learning algorithms.

- Text cleaning involves removing noise, irrelevant information, and unwanted characters from the text.
- Tokenization is the process of splitting text into individual tokens or words. It is a fundamental step that forms the basis for further analysis.
- Stop words are common words that do not carry significant meaning for hate speech detection. These words, such as "and," "the," or "is," can be removed to reduce noise and improve processing efficiency.
- Stemming and lemmatization are techniques to reduce words to their base or root forms. These techniques help to normalize the text and reduce feature dimensionality.
- Feature extraction involves converting text data into a numerical representation that machine learning algorithms can process.

These preprocessing steps help in standardizing and transforming the raw text into a structured format that can be fed into machine learning algorithms. The specific steps and techniques used may vary depending on the requirements of the hate speech detection task, the characteristics of the dataset, and the choice of machine learning models.

C. Machine Learning Techniques:

Here are the details on how each machine learning technique, namely Support Vector Machines (SVM), Random Forest, and Logistic Regression, can be implemented and trained for hate speech detection:

1. *Support Vector Machines (SVM):*
    - Implementation: SVM is implemented by using a suitable machine learning library or framework that supports SVM, such as scikit-learn in Python.
    - Data Preparation: The hate speech dataset is preprocessed, including text cleaning, tokenization, and feature extraction steps as discussed earlier.
    - Feature Representation: The preprocessed text data is transformed into numerical feature vectors, such as Bag-of-Words (BoW), TF-IDF, or word embeddings.
    - Model Training: The SVM model is trained using the preprocessed feature vectors and their corresponding labels (hate speech or non-hate speech). The SVM algorithm seeks to find the optimal hyperplane that separates the two classes with the largest margin.
    - Hyperparameter Tuning: The model's performance can be further optimized by tuning SVM-specific hyperparameters, such as the choice of kernel (linear, polynomial, radial basis function) and the regularization parameter (C).
    - Evaluation: The trained SVM model is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score on a separate test dataset. Cross-validation can also be employed for more robust evaluation.
2. *Random Forest:*
    - Implementation: Random Forest can be implemented using libraries like scikit-learn in Python, which provide Random Forest classifiers.
    - Data Preparation: Similar to SVM, the hate speech dataset undergoes preprocessing steps such as text cleaning, tokenization, and feature extraction.
    - Feature Representation: The preprocessed text data is transformed into numerical feature vectors using techniques like BoW, TF-IDF, or word embeddings.
    - Model Training: Random Forest is trained by building an ensemble of decision trees. Each decision tree is trained on a randomly sampled subset of the data with replacement (bootstrapping) and using a random subset

of features. The predictions from individual trees are combined through voting or averaging to make the final prediction.

- Hyperparameter Tuning: The performance of the Random Forest model can be optimized by tuning hyperparameters like the number of trees, maximum depth of trees, and the number of features considered for each split.
- Evaluation: The trained Random Forest model is evaluated using appropriate evaluation metrics on a separate test dataset, and cross-validation can also be applied for robust evaluation.

3. *Logistic Regression:*
    - Implementation: Logistic Regression can be implemented using libraries like scikit-learn in Python, which provide Logistic Regression classifiers.
    - Data Preparation: The hate speech dataset undergoes text preprocessing steps such as text cleaning, tokenization, and feature extraction.
    - Feature Representation: The preprocessed text data is transformed into numerical feature vectors using techniques like BoW, TF-IDF, or word embeddings.
    - Model Training: Logistic Regression models the probability of an instance belonging to a particular class using a logistic function. The model is trained by optimizing the parameters (weights and biases) to maximize the likelihood of the observed labels given the features using optimization techniques like gradient descent.
    - Hyperparameter Tuning: Hyperparameters like the regularization parameter (C) controlling the trade-off between model complexity and overfitting can be tuned to optimize the performance of the Logistic Regression model.
    - Evaluation: The trained Logistic Regression model is evaluated using appropriate evaluation metrics on a separate test dataset, and cross-validation can also be used for reliable evaluation.

In all these techniques, it is essential to split the dataset into training and testing subsets to ensure unbiased evaluation. The models are trained on the training subset and evaluated on the testing subset using suitable metrics to assess their performance in hate speech detection. Additionally, hyperparameter tuning and cross-validation techniques can be applied to optimize and validate the models' performance.

D. Experimental Setup

In this step, a test set is used to evaluate the performance of the built-in classifier on unlabelled textual data, specifically for categorizing content into "hate speech" or "clean speech" categories. The effectiveness of the classifier can be analyzed by computing the true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP), which together form the confusion matrix.

To evaluate the overall performance of the developed classifier, several common performance measures for text content categorization are commonly used. Here are brief explanations of some key performance indicators:

Precision:

- Precision represents the percentage of predicted positives that are actually positive.
- It can be calculated as TP divided by the sum of TP and FP.
- Alternative definitions include the number of accurate positive outputs the model produced or the proportion of correctly anticipated positive classes that actually materialized.

Recall:

- Recall represents the proportion of actual positive instances that are correctly identified by the model.
- It can be calculated as TP divided by the sum of TP and FN.
- Recall measures the model's ability to identify all relevant instances.

F-Measure:

- The F-measure is the harmonic mean of precision and recall.
- It provides a balanced evaluation of precision and recall, giving equal weight to both metrics.
- By considering both precision and recall, the F-measure is useful for comparing models that have different trade-offs between these two metrics.
- The F-score is highest when precision and recall are equal.

Accuracy:

- Accuracy is an important consideration for evaluating the overall accuracy of a classification problem.
- It measures how often the model correctly predicts the outcome.
- It can be calculated by dividing the total number of correct predictions (TP and TN) by the total number of predictions made by the classifier.

The confusion matrix includes True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). TP represents the correctly identified positive instances (hate speech), FP represents non-hate speech incorrectly classified as hate speech, FN represents hate speech incorrectly classified as non-hate speech, and TN represents correctly identified non-hate speech.

By analyzing these performance measures, such as precision, recall, F-measure, and accuracy, we can evaluate the effectiveness and overall performance of the developed classifier in detecting and categorizing bogus news accurately in the hybrid data environment.

5. RESULTS AND DISCUSSIONS

The generated data is sourced from English datasets obtained from GitHub and other social media websites. The data is divided into a testing set, which constitutes 35% of the data, and a training set, which comprises the remaining 65%. Prior to training, pre-processing activities are performed on the data to prepare it for analysis. Various machine learning techniques are employed, including Support Vector Machines (SVM), logistic regression, and random forest. Each technique is trained on the training data and then evaluated using performance evaluation parameters. The confusion matrix is utilized to facilitate this evaluation. The confusion matrix provides a comprehensive view of a classification model's performance for a given set of test data. The order of the matrix is determined by the number of classes in the problem. The rows of the matrix represent the examples in each actual class, while the columns (or vice versa) represent the examples in each predicted class. The confusion matrix allows for a comparison between the predicted output of each model and the actual class of the data. This comparison helps assess the accuracy, precision, recall, and other performance metrics of each machine learning model. By analyzing the confusion matrix, the effectiveness of SVM, logistic regression, and random forest in classifying the test data can be determined.

Table1: Comparative study of SVM, Logistic and Random Forest

SPECIFICATIONS	SVM	LOGISTIC REGRESSION	RANDOM FOREST
ACCURACY	92	83	88
PRECISION	93	87	88
RECALL	91	82	86
F1-SCORE	92	81	87

Furthermore, the study's findings may differ based on the evaluation metrics employed and the specific research objectives. Evaluating hate speech detection models using multiple metrics such as accuracy, precision, recall, and F1-score provides a comprehensive understanding of their performance and helps assess their suitability for real-world applications.

6. CONCLUSION

In conclusion, Support Vector Machines (SVM), Random Forest, and Logistic Regression have demonstrated their potential for hate speech detection on social media platforms. Each technique offers unique strengths and characteristics that contribute to their effectiveness in identifying hate speech instances. SVM, with its ability to handle high-dimensional feature spaces and non-linear decision boundaries, has shown competitive performance in distinguishing hate speech from non-hate speech content. Its generalization ability and robustness against overfitting make it a reliable choice for hate speech detection tasks. Random Forest, through its ensemble of decision trees, has proven effective in handling complex feature interactions and non-linear relationships. Its ability to capture important features and patterns in text data has contributed to successful hate speech detection. Logistic Regression, despite its simplicity, has demonstrated effectiveness in hate speech detection. Its interpretability and efficiency make it an attractive choice, particularly when the focus is on modeling the probability of hate speech occurrence.

While these techniques have shown promise, it is important to note that hate speech detection is a challenging and evolving task. The effectiveness of SVM, Random Forest, and Logistic Regression can be influenced by factors such as dataset characteristics, feature representation, and hyperparameter tuning. Additionally, advancements in deep learning models and transformer-based architectures should also be considered for hate speech detection tasks.

In summary, SVM, Random Forest, and Logistic Regression have demonstrated their potential in hate speech detection. By leveraging their unique characteristics, researchers and practitioners can develop effective and tailored solutions to combat hate speech on social media platforms, fostering safer and more inclusive online environments.

7. REFERENCES

1. Granik Mykhailo and Mesyura Volodymyr 2017 First Ukraine Conference on Electrical and Computer Engineering (UKRCON) (Ukraine: IEEE) Fake news detection using naive Bayes classifier.

2. Gilda S. 2017 15th Student Conference on Research and Development (SCOReD) (IEEE) Evaluating machine learning algorithms for fake news detection 110-115.

3. Jain Akshay and Kasbe Amey 2018 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (Bhopal, India: IEEE) Fake News Detection.

4. Aphiwongsophon Supanya et al 2018 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (Chiang Rai, Thailand, Thailand: IEEE) Detecting Fake News with Machine Learning Method.

5. Sharma K., Qian F., Jiang H., Ruchansky N., Zhang M. and Liu Y. 2019 Combating fake news: A survey on identification and mitigation techniques ACM Transactions on Intelligent Systems and Technology (TIST) **10** 1-42.

6. Zhang, Jiawei, Dong Bowen and Yu Philip S. 2020 2020 IEEE 36th International Conference on Data Engineering (ICDE) (IEEE) Fakedetector: Effective fake news detection with deep diffusive neural network.

7. N. Hoy and T. Koulouri, “A systematic review on the detectionof fake news articles,”2021, http://arxiv.org/abs/2110.11240.

8. R. D. Abdiansyah, D. Mutiara, S. P. Sumedha, andN. Hanafiah, “Effective methods for fake news detection: a sys-tematic literature review,”in 2021 1st International Conferenceon Computer Science and Artificial Intelligence (ICCSAI), vol.1, pp. 278–283, Jakarta, Indonesia, 2021

9. M. Albahar, “A hybrid model for fake news detection: leverag-ing news content and user comments in fake news,”IET Information Security, vol. 15, no. 2, pp. 169–177, 2021.

10. M. Celliers and M. Hattingh, A Systematic Review on FakeNews Themes Reported in Literature, vol. 12067, SpringerInternational Publishing, LNCS, 2020.

11. T. Chauhan and H. Palivela, “Optimization and improvementof fake news detection using deep learning approaches forsocietal benefit,”International Journal of Information Man-agement Data Insights, vol. 1, no. 2, article 100051, 2021.

12. S. Deepak and B. Chitturi, “Deep neural approach to fake-news identification,”Procedia Computer Science, vol. 167,no. 2019, pp. 2236–2243, 2020.

13. Drif and S. Giordano, “Fake news detection method basedon text-features,”in International Academy, Research, andIndustry Association (IARIA), vol. 23no. 3, pp. 26–31, France,2019.

14. M. K. Elhadad, K. Fun Li, and F. Gebali, “Fake news detectionon social media: a systematic survey,”in 2019 IEEE Pacific RimConference on Communications, Computers and Signal Pro-cessing (PACRIM), Victoria, BC, Canada, 2019.

15. A. A. Ahmed, A. Aljarbouh, P. K. Donepudi, and M. S.Choi, “Detecting fake news using machine learning: a system-atic literature review,”Journal of Educational Psychology,vol. 58, no. 1, pp. 1932–1939, 2021.

16. D. de Beer and M. Matthee, Approaches to Identify Fake News:A Systematic Literature Review, vol. 136, no. Macaulay 2018,2021Springer International Publishing, 2021.

17. F. D. C. Medeiros and R. B. Braga, “Fake news detection insocial media: a systematic review,”The ACM InternationalConference Proceeding Series, vol. 3, no. 5, pp. 2–7, 2020.

18. D’Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, “Fakenews detection: a survey of evaluation datasets,”Computer Sci-ence - PeerJ, vol. 7, no. 2, pp. e518–e534, 2021.

19. Pilkevych, D. Fedorchuk, O. Naumchak, andM. Romanchuk, “Fake news detection in the framework ofdecision-making system through graph neural network,”in2021 IEEE 4th International Conference on Advanced Informa-tion and Communication Technologies (AICT), pp. 153–157,Lviv, Ukraine, 2021.

20. S. Preston, A. Anderson, D. J. Robertson, M. P. Shephard, andN. Huhe, “Detecting fake news on Facebook: the role of emo-tional intelligence,”PLoS One, vol. 16, no. 3, pp. 1–13, 2021.

21. Girgis, S., Amer, E., Gadallah, M.: Deep learning algorithms for detecting fake news in online text. IEEE (2018). 978-1-5386-5111-7/18/$31.00

22. Lillie, A.E., Middelboe, E.R.: Fake news detection using stance classification: a survey. arXiv:1907.00181v1 [cs.CL]. Accessed 29 June 2019

23. Ajao, O., Bhowmik, D., Zargari, S.: Sentiment aware fake news detection on online social networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019). https://doi.org/10.1109/icassp.2019.8683170.

24. Li, S., Ma, K., Niu, X., Wang, Y., Ji, K., Yu, Z., Chen, Z.: Stacking-based ensemble learning on low dimensional features for fake news detection. In: 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (2019). https://doi.org/10.1109/hpcc/smartcity/dss.2019.00383

25. Mahabub, A.: A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers. SN Appl. Sci. 2, 525 (2020). https://doi.org/10.1007/s42452-020-2326-y.

26. Reddy, P.B.P., Reddy, M.P.K., Reddy, G.V.M., Mehata, K.M.: Fake data analysis and detection using ensembled hybrid algorithm. In: 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (2019). https://doi.org/10.1109/iccmc.2019.8819741

27. N. Hoy and T. Koulouri, “A systematic review on the detection of fake news articles,” 2021, http://arxiv.org/abs/2110.11240.

28. R. D. Abdiansyah, D. Mutiara, S. P. Sumedha, and N. Hanafiah, “Effective methods for fake news detection: a systematic literature review,” in 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), vol. 1, pp. 278–283, Jakarta, Indonesia, 2021.

29. O. Abu Arqoub, A. Abdulateef Elega, B. Efe Özad, H. Dwikat, and F. Adedamola Oloyede, “Mapping the scholarship of fake news research: a systematic review,” Journalism Practice, vol. 16, no. 1, pp. 56–86, 2022.

30. T. Chauhan and H. Palivela, “Optimization and improvement of fake news detection using deep learning approaches for societal benefit,” International Journal of Information Management Data Insights, vol. 1, no. 2, article 100051, 2021.

31. B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, “Trends in combating fake news on social media – a survey,” Journal of Information and Telecommunication, vol. 5, no. 2, pp. 247–266, 2021.

32. F. C. D. da Silva, R. V. da Costa Alves, and A. C. B. Garcia, “Can machines learn to detect fake news? A survey focused on social media,” in Hawaii International Conference on System Sciences (HICSS), vol. 2019, pp. 2763–2770, Grand Wailea, Hawaii, 2019.

33. S. Deepak and B. Chitturi, “Deep neural approach to fake-news identification,” Procedia Computer Science, vol. 167, no. 2019, pp. 2236–2243, 2020.

34. S. Khan, S. Hakak, N. Deepa, B. Prabadevi, K. Dev, and S. Trelova, “Detecting COVID-19-related fake news using feature extraction,” Frontiers in Public Health, vol. 9, no. January, pp. 1–9, 2021.

35. R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. Ali Almazroi, “A fine-tuned BERT-based transfer learning approach for text classification,” Journal of healthcare engineering, vol. 2022, Article ID 3498123, 17 pages, 2022.

36. Drif and S. Giordano, “Fake news detection method based on text-features,” in International Academy, Research, and Industry Association (IARIA), vol. 23, no. 3, pp. 26–31, France, 2019.

37. A.A. A. Ahmed, A. Aljarbouh, P. K. Donepudi, and M. S. Choi, “Detecting fake news using machine learning: a systematic literature review,” Journal of Educational Psychology, vol. 58, no. 1, pp. 1932–1939, 2021.

38. O. D. Apuke and B. Omar, “Fake news and COVID-19: modelling the predictors of fake news sharing among social media users,” Telematics and Informatics, vol. 56, article 101475, 2021.

39. Guimarães, Á. Figueira, and L. Torgo, “Can fake news detection models maintain the performance through time? A longitudinal evaluation of twitter publications,” Mathematics, vol. 9, no. 22, 2021.

40. Machete and M. Turpin, The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review, Springer International Publishing, LNCS, vol. 12067, 2020.

41. C. Melchior and M. Oliveira, “Health-related fake news on social media platforms: a systematic literature review,” New Media & Society, vol. 1, no. 23, 2021.

42. C. V. Meneses Silva, R. Silva Fontes, and M. Colaço Júnior, “Intelligent fake news detection: a systematic mapping,” Journal of Applied Security Research, vol. 16, no. 2, pp. 168–189, 2021.

43. M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, “Detection of bots in social media: a systematic review,” Information Processing & Management, vol. 57, no. 4, 2020.

44. I.Segura-Bedmar and S. Alonso-Bartolome, “Multimodal fake news detection,” Information, vol. 13, no. 6, p. 284, 2022.

45. W. S. Paka, R. Bansal, A. Kaushik, S. Sengupta, and T. Chakraborty, “Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection,” Applied Soft Computing, vol. 107, article 107393, 2021.

46. I.Pilkevych, D. Fedorchuk, O. Naumchak, and M. Romanchuk, “Fake news detection in the framework of decision-making system through graph neural network,” in 2021 IEEE 4th International Conference on Advanced Information and Communication Technologies (AICT), pp. 153–157, Lviv, Ukraine, 2021.

47. S. Preston, A. Anderson, D. J. Robertson, M. P. Shephard, and N. Huhe, “Detecting fake news on Facebook: the role of emotional intelligence,” PLoS One, vol. 16, no. 3, pp. 1–13, 2021.

48. A.Reyes-Menendez, J. R. Saura, and F. Filipe, “The importance of behavioral data to identify online fake reviews for tourism businesses: a systematic review,” PeerJ Computer Science, vol. 5, no. 9, pp. 1–21, 2019.

49. S. Shahin, B. Ang, and N. D. Anwar, Disinformation and fake news, Journal Mass Commun, 2022.

50. I.Shu, H. R. Bernard, and H. Liu, “Studying fake news via network analysis: detection and mitigation,” Summer Tutor, vol. 3, no. 5, pp. 43–65, 2019.

51. I. Stahl, “Fake news detector in online social media,” International Journal of Engineering and Advanced Technology, vol. 9, no. 1S4, pp. 58–60, 2019.

52. X. Zhang and A. A. Ghorbani, “An overview of online fake news: characterization, detection, and discussion,” Information Processing and Management, vol. 57, no. 2, article 102025, 2020.

53. G. Xu and H. Jin, “Using artificial intelligence technology to solve the electronic health service by processing the online case information,” Journal of Healthcare Engineering, vol. 2021, Article ID 9637018, 12 pages, 2021.

54. S. L. Ting, W. H. Ip, and A. H. C. Tsang, “Is Naïve bayes a good classifier for document classification?” International Journal of Software Engineering and Its Applications, vol. 5, no. 3, pp. 37–46, 2021.

55. S. Karthika and N. Sairam, “A Naïve Bayesian classifier for educational qualification,” indian journal of science and technology, vol. 8, no. 16, 2021.

56. W. Dai, G. R. Xue, Q. Yang, and Y. Yu, “Transferring naive Bayes classifiers for text classification,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 1, pp. 540–545, 2022.

57. I. Sánchez-Torné, J. C. Morán-Álvarez, and J. A. Pérez-López, “The importance of corporate social responsibility in achieving high corporate reputation,” Corporate Social Responsibility and Environmental Management, vol. 27, no. 6, pp. 2692–2700, 2020.

58. X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, “Fake news early detection: an interdisciplinary study,” pp. 3207-3208, 2019, http://arxiv.org/abs/1904.11679.

59. Y. Wang and J. Y. Xu, “An autonomous semantic learning methodology for fake news recognition,” in 2021 IEEE International Conference on Autonomous Systems (ICAS), pp. 1–6, Montreal, QC, Canada, 2021.

60. Z. Liang, J. Liu, A. Ou, H. Zhang, Z. Li, and J. X. Huang, “Deep generative learning for automated EHR diagnosis of traditional Chinese medicine,” Computer Methods and Programs in Biomedicine, vol. 174, pp. 17–23, 2019.

61. M. Basaldella, F. Liu, E. Shareghi, and N. Collier, "COMETA: a corpus for medical entity linking in the social media," in EMNLP 2020. 2020 Conference on Empirical Methods in. Natural Language Processing, vol. 2, no. 1, pp. 3122–3137, 2020.