Research Article

Using and Comparison of Artificial Intelligence Techniques to Detect Misinformation and Disinformation on Twitter

Omar Raad Mahmood Mahmood[1*], Funda Akar[2]

[1] Erzincan Binali Yıldırım University, Institute of Science and Technology, Orcid ID: https://orcid.org/0009-0009-4831-2066, E-mail: omar18.or@gmail.com
[2] Erzincan Binali Yıldırım University, Department of Computer Engineering, Orcid ID: https://orcid.org/000-0001-9376-8710, E-mail: fakar@erzincan.edu.tr
[*] Correspondence: omar18.or@gmail.com

Received 21 January 2023
Received in revised form 12 March 2024
In final form 26 May 2024

Reference: Mahmood, O., R., M., Akar, F. Using and Comparison of Artificial Intelligence Techniques to Detect Misinformation and Disinformation on Twitter. The European Journal of Research and Development, 4(2), 254-264.

Abstract

This research investigates diverse artificial intelligence (AI) techniques for detecting misinformation on Twitter, addressing the pervasive concern of misinformation and fake news affecting public discourse. Employing models such as Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Random Forest Classifier, Multinomial Naive Bayes and Gradient Boosting Classifier, we discern deceptive content from reliable information. Utilizing a dataset of 23,481 false tweets and approximately 21,417 real tweets, our analysis leverages Natural Language Processing (NLP), Deep Learning (DL) and Machine Learning (ML) techniques, showcasing the effectiveness of each model in identifying misinformation patterns. Our investigation rigorously assesses the strengths and limitations of AI techniques, focusing on accuracy, efficiency and scalability. Notably, the best results are achieved by models such as LSTM (98.84% accuracy, 98.79% F1 score), SVM (99.44% accuracy, 99.44% F1 score) and XGBoost Classifier (99.82% accuracy, 99.81% F1 score). The findings provide valuable insights into the performance of key models and serve as a resource for academics and researchers in the fields of artificial intelligence and social media analysis. Additionally, they provide practical guidance for supporting information integrity on Twitter, contributing to ongoing efforts to combat misinformation and enhance information credibility.

Keywords: *Misinformation detection, artificial intelligence, twitter, AI techniques, social media.*

1. Introduction

In the era of digital transformation, social media has become an integral part of daily life, providing diverse news sources and direct communication without borders. Anyone can publish fake news on these social media platforms at a very low cost [1]. The Pew Research Center indicates that in 2018, 68% of American adults got their news from social media, demonstrating the important role these platforms play in disseminating information [2]. Social networking sites play a crucial role in facilitating virtual interactions, enabling the real-time exchange of ideas and news on various issues. Twitter, in particular, has emerged as an important outlet, generating interactions and influencing public conversations. This is due to the core properties of social networks, namely their wide accessibility, ease of use and rapid spread, which can further magnify the impact of fake news [3, 4]. Misinformation, especially in the form of fake news and rumors, is a major challenge on online social networks such as Facebook, Twitter and Sina Weibo. For example, the OTT platform's misinformation, "We'll give you three months of Netflix Premium to help you get through this," Time at home due to the outbreak. Coronavirus (Covid-19)" [5].

Fake news detection is a current area of research that can be approached from multiple disciplines [6, 7]. In recent years, many attempts have been made to distinguish between fake news and real news [8]. The emergence of AI represents a new era in the battle against deceptive content. The research explores how cutting-edge AI techniques, such as ML algorithms, NLP and DL models, can be applied to distinguish real information from deceptive content. The comparative aspect of this study provides a nuanced dimension, assessing the strengths and weaknesses of different AI approaches in the context of misinformation detection. The algorithm's accuracy depends on various factors, such as the quality of the training data, the complexity of the news story and the sophistication of the fake news creator [9]. By examining performance metrics, accuracy rates and computational efficiency, the research aims to contribute valuable insights to the ongoing discourse about fortifying social media platforms against the insidious influence of misinformation. It provides a promising set of techniques to not only detect but also compare and evaluate different strategies aimed at identifying misinformation on Twitter.

This study comprehensively evaluates current research on misinformation detection based on ML, DL and NLP methods, with the aim of understanding the landscape of current challenges, solutions and validations. Especially since artificial intelligence (AI) algorithms offer a promising arsenal of techniques to not only detect but also compare and evaluate the effectiveness of different strategies in identifying misinformation on Twitter. While previous studies have addressed similar topics, our work is distinctive in several key respects.

A comprehensive comparison was conducted using a variety of AI models, including LSTM, SVM, Random Forest, Multinomial Naive Bayes and Gradient Boosting Classifier, to assess their effectiveness in detecting misinformation. A dataset comprising a substantial number of false tweets (23,481) and authentic tweets (approximately 21,417) was employed to ensure a robust evaluation of the models' performance across a diverse range of content. Furthermore, our approach integrates diverse techniques from NLP, DL and ML.

2. Literature reviews

Several studies have significantly contributed to the advancement of fake news detection systems, showcasing the effectiveness of AI algorithms. In a groundbreaking work by [10], a hybrid DL model integrates convolutional and RNN, augmented by generative adversarial networks, achieving an impressive accuracy of 93.87%. Similarly, [11] employs n-gram analysis and ML algorithms, demonstrating that fine-tuning with Random Search CV enhances accuracy to 94.2%. Other researchers [12] explore the integration of SVM with Naïve Bayes and NLP, resulting in a remarkable accuracy of 93.6%, particularly effective in regions like India. In the realm of social media, [13] introduces a hybrid model, CNN-LSTM-SVM, combining CNN, LSTM and SVM to attain an average score of 82.55% in five-fold cross-validation. The researchers [14] employ CNN and RNN-LSTM with Word2vec embeddings, achieving superior accuracy rates in English and Turkish language systems. Meanwhile, [15] proposes a CNN-RNN hybrid model, achieving an accuracy of 95.12% for hoax detection. study [16] proposed an ensemble-based deep learning solution. The strategy used two essential models: a Bi-LSTM-GRU-dense model for textual analysis and a dense deep learning model for additional attributes, based on the LIAR dataset. The study achieved an accuracy of 0.898, with a recall of 0.916, precision of 0.913 and F-score of 0.914, demonstrating significant progress in fake news detection compared to previous works. Researchers [17] delve into attention-based neural networks, with the word-level memory network outperforming state-of-the-art models by 8.4% and 4.9% in development and test sets, respectively. The WELFake model achieves a high accuracy rate of 96.73% in detecting fake news, surpassing the BERT and CNN models by 1.31% and 4.25%, respectively. It combines linguistic features with word embedding and employs SVM as the most accurate machine learning model, showcasing the efficacy of feature selection and ensemble voting classification [18]. The identification of false information is examined in different linguistic contexts. For instance, [19] has developed a deep learning model using MARBERT with CNN architecture to identify fake news in Arabic. This algorithm has achieved remarkable accuracy and an F1 score of 0.956, surpassing other methods. The study aimed to tackle the challenge of spreading fake news in Amharic, a language with limited resources, during the era of new media. The researchers [20] obtained a new dataset of 12,000 annotated news articles from Facebook and implemented DL algorithms like Bi-GRU and CNN for detecting fake news automatically. The CNN model exhibited excellent performance, achieving an accuracy of 93.92% and an f1-score of 94%. The study [21] evaluated the performance of several false news detection algorithms, including SVM, Naive Bayes, RNN, and CNN, using the ISOT dataset. CNN achieved the best results with 92% accuracy, 90% precision, 88% recall and a 90% F1 score on the validation set. In a distinctive approach, [22] proposes a hybrid model utilizing two-way LSTM, metadata and rumor path diffusion, attaining an accuracy of 94.3%. This model outperforms contemporary solutions for discerning between real and fake news based on various characteristics. A combined machine learning approach, incorporating algorithms and text features extracted using the LIWC tool, is explored by [23], achieving an impressive accuracy of 99% on DS1 with Random Forest and Perez-LSVM. Lastly, [24] utilizes the "2010 Chile Earthquake Dataset" from Twitter, incorporating feature engineering to enhance accuracy. SVM and Naïve Bayes emerge as superior algorithms, underscoring the significance of processing textual data using Count Vectors and TF-IDF in identifying fake news in Twitter threads. The research employs traditional machine learning models for fake news detection, utilizing supervised learning algorithms. The XGBoost algorithm shows the highest accuracy of over 75%, followed by SVM and Random Forest with approximately 73% accuracy [25]. The research [26] addresses the detection of fake news using linguistic feature extraction and sequential neural network models and syntactic, grammatical, sentimental and readability features. The combined linguistic feature-based sequential neural network achieved 86% classification accuracy. This approach outperformed initial attempts using ML algorithms, which achieved a maximum accuracy of 72% using ensemble methods. This additional study [27] investigates the detection of fake news on social media, focusing solely on textual features using stylometric analysis and text-based word vector representations. Ensemble methods such as boosting yield the most promising results. With an accuracy of up to 95.49%, the combination of stylometric and word vector features proves effective, demonstrating the importance of these methods in achieving high accuracy without relying on metadata or user-related information.

This comprehensive literature review underscores the richness and diversity of methodologies employed in the detection of fake news, providing a robust foundation for the development of our research in this crucial domain. The amalgamation of AI algorithms, linguistic analysis and innovative models demonstrates the evolving landscape of fake news detection, inviting further exploration and refinement.

3. Materials and Methods

The chapter outlines the research objectives, tools, techniques, data acquisition, preparation & annotation, pre-processing, data representation, model building, training procedures and typical evaluation methods. Flowchart of the work is shown in Figure 1.

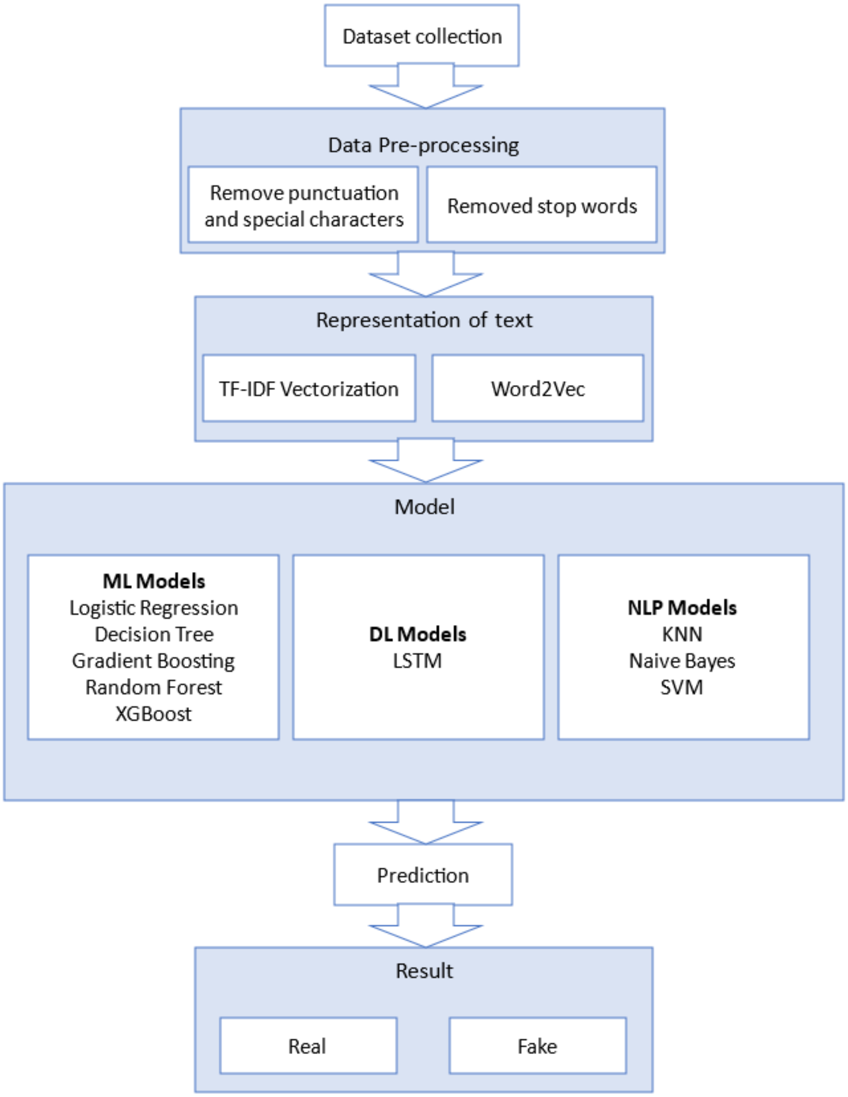

Figure 1: Flowchart of the work

2.1. Dataset

In this research, we utilized a comprehensive dataset comprising two distinct files [28]. The IISOT dataset comprises 44,898 news statements, sourced from Reuters.com on April 5, 2022. The authentic news was gathered from reliable sources, while the misleading information was obtained from untrustworthy sources. Of these, 21,417 are authentic and 23,481 are fake [29],[45]. Many researchers, including [30–33], also used the ISOT dataset.

A: Real News Dataset: The Real News dataset is a compilation of 21,417 articles categorized as authentic news. These articles cover a spectrum of topics, with 53% dedicated to global politics and the remaining 47% encompassing various other worldrelated subjects. The publication dates of the news span from January 13, 2016, to December 31, 2017.

B: Fake News Dataset: The fake news dataset comprises 23,481 articles identified as spurious news sources. Among these, 39% pertain to general news, 29% focus on political content and the remaining 32% cover miscellaneous topics. The publication timeline for these fake news articles extends from March 31, 2015, to February 19, 2018.

2.2. Data Pre-processing

Text preprocessing using a stopword list in order to classify news articles. The text was converted to lower case, removed non-word characters and punctuation marks. Machine learning models were trained to identify patterns and predict classification. The TF-IDF step transformed the raw text into a digital format and each model's performance was evaluated.

2.3. Representation of text

The raw text data is converted into a matrix using a TF-IDF vector where each row corresponds to a document and each column corresponds to a unique term.
Word2Vec is used to generate word embeddings from a set of text data. Word2Vec learns distributed representations of words in a continuous vector space by predicting context words given a target word or vice versa. The resulting word embeddings capture semantic relationships between words and can be used as features in various natural language processing tasks.

2.4. Artificial intelligence models

In the pursuit of developing robust methodologies for discerning authentic and deceptive news articles, real-time detection of autonomously generated fake news is a significant challenge for machine learning and natural language processing [34]. A suite of artificial intelligence algorithms has been applied. Each algorithm is meticulously designed to address the intricate challenges posed by the task of detecting fake news. The models considered in this research span a spectrum of ML, DL and NLP techniques, each contributing a unique perspective to the overarching goal of accurate classification.

- **Logistic Regeresion Classifier of machine learning:** The logistic regression (LR) and term frequency-inverse document frequency (TF-IDF) method is a reliable method for distinguishing between genuine and deceptive news articles. TF-IDF transforms textual documents into numerical vectors, summarizing the importance of individual terms. LR uses these vectors for classification, estimating the weights for each term's TF-IDF score. The final decision boundary is determined through the linear combination of weighted features, providing an interpretable framework for news classification.

- **Decision Tree Classifier of machine learning:** Decision trees are a data classification algorithm used to distinguish between authentic and fabricated news content using TF-IDF. Each document is treated as a set of words and weights are assigned based on their importance. TF-IDF quantifies the significance of a word within a document based on its prevalence across all documents. The decision tree partitions the dataset, making decisions based on TF-IDF values and identifying key terms for news article classification. The product of TF and IDF is the TF-IDF score. At each decision node, the algorithm selects the optimal feature and threshold for data splitting based on class labels.
- **Gradient Boosting Classifier of machine learning:** The Gradient Boosting classifier, an ensemble learning algorithm, combines weak learners to create robust prediction models in the classification of real and fake news, utilising the TF-IDF technique. The weak learners are designed to iteratively correct previous errors by adapting to the residuals of the combined predictions. This process is intended to minimise a given loss function in order to improve the overall prediction accuracy. The scikit-learn implementation with a fixed seed ensures the reproducibility of the results. In the TF-IDF representation, the algorithm utilises a feature matrix, wherein each row corresponds to a document and each column represents a unique word. The TF-IDF algorithm evaluates the importance of words and is able to distinguish between real and fake news based on these weighted features.
- **Random forest classifier of machine learning:** The random forest classifier, an ensemble learning algorithm that builds decision tree forests, classifies real and fake news using TF-IDF. The random forest (random_state=5) reduces overfitting and increases robustness by generating trees trained on random subsets of data with replacement. The random_state parameter ensures repeatability. In TF-IDF, the algorithm uses numerical values to capture word importance and discriminates between real and fake articles. Multiple decision trees make decisions based on different subsets of features and instances, and predictions are aggregated by majority voting. The computations involve building the TF-IDF matrix, randomly selecting subsets, training the trees and collecting the outputs for robust and accurate classification.
- **LSTM classifier of deep learninig:** This research paper presents a DL model for classifying real and fake news articles using Word2Vec embeddings and long short-term memory (LSTM) networks. The model is divided into training and testing sets, with 20% for testing and 80% for training. The model is trained using the fit method with the X_train and y_train datasets, with 30% for validation. The model consists of an embedding layer, an LSTM layer and a dense layer. The embedding layer

transforms input words into dense vectors, while the LSTM layer captures sequential patterns and dependencies. The dense layer outputs a probability score, indicating the likelihood of an article being real or fake. The model is suitable for binary classification tasks involving text data.

- **k-Nearest Neighbors classifier of natural languge processing:** The provided algorithm is a K-Nearest Neighbours (KNN) classifier, an instance of ML methodology utilised for the classification of text data. It incorporates Word2Vec embeddings, a WordNet lemmatizer for lemmatisation and a CountVectorizer for feature extraction. The resultant feature vectors are trained to differentiate between news articles deemed authentic and those identified as false based on the textual content of the articles. The algorithm computes distances between feature vectors and determines class labels based on a majority vote among the K-nearest neighbours of the feature vectors. The architecture comprises preprocessing steps, the KNN model, and training and testing processes.
- **Naive Bayes classifier of natural languge processing:** The Naive Bayes classifier is a probabilistic class technique based on Bayes' theorem, used in spam filtering, sentiment analysis and record categorization. It is particularly effective for text classes and is used in spam filtering, sentiment analysis and record categorization. In fake news detection, the Naive Bayes algorithm uses the frequencies of words in documents to calculate the likelihood of a document belonging to a particular class. The code uses natural language processing techniques, such as CountVectorizer and Word2Vec and WordNetLemmatizer, to improve the model's performance.
- **Support Vector Machine classifier of natural languge processing:**The algorithm is a SVM classifier, implemented using the scikit-learn library in Python, used for fake news detection. It uses a training and testing procedure to identify genuine and deceptive news articles. The model is trained using WordNetLemmatizer, Word2Vec, and CountVectorizer, which enhance its understanding of language. The combination of these techniques improves the model's ability to discern linguistic nuances and distinguish between trustworthy and deceptive information. The research presents various models, each with its own strengths, contributing to the advancement of fake news detection methodologies.

4. Results

The results obtained from the experimentation with various machine learning and deep learning models for the classification of real and fake news are in Table 1. The models include DL models such as LSTM, NLP models like GaussianNB, KNeighbors, SVM and ML models comprising Logistic Regression, Decision Tree, Gradient Boosting, Random Forest, and XGBoost.

Following an analysis of the performance metrics, it was found that the XGBoost model exhibited the highest accuracy (99.82%) and F1 score (99.82%), while the GaussianNB model displayed the lowest accuracy (93.38%) and F1 score (93.28%). The results provide evidence that machine learning models are effective in distinguishing between authentic and fabricated news articles.

Table 1: A summary of the success rates of AI models

	Model	Accuracy	F1-score
ML methods	XGBoost	**99.82%**	**99.81%**
	logistic regression	98.82%	98.76%
	decision tree	99.60%	99.58%
	gradient boosting	99.63%	99.61%
	Random forest	99.00%	98.94%
NLP methods	SVM	99.44%	99.44%
	knn	74.81%	79.11%
	GaussianNB	93.38%	93.28%
DL methods	LSTM	98.84%	98.79%

5. Discussion and Conclusion

This study evaluated various artificial intelligence techniques for detecting misinformation on Twitter. Using models like LSTM, SVM, Random Forest Classifier, k-NN, Multinomial Naive Bayes and Gradient Boosting Classifier, the study aimed to distinguish deceptive content from reliable information. The research assessed the strengths and limitations of each AI model using Twitter data and natural language processing. The findings revealed the effectiveness of these models in identifying misinformation patterns and false narrative propagation. For example, LSTM (98.84% accuracy, 98.79% F1 score), SVM (99.44% accuracy, 99.44% F1 score), and XGBoost Classifier (99.82% accuracy, 99.81% F1 score). The study also highlighted the importance of privacy, bias mitigation, and freedom of expression in AI-based misinformation detection. The findings provide valuable insights for academics, researchers and practitioners in AI and social media analysis, as well as practical guidance for social media companies and policymakers.

References

[1] Song, C., Ning, N., Zhang, Y. & Wu, B. (2021). A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Inf process manag 58.

[2] Shu, K., Cui, L., Wang, S., Lee, D. & Liu, H. (2019). Defend: explainable fake news detection. In proceedings of the acm sigkdd international conference on knowledge discovery and data mining pp 395–405. Association for computing machinery.

[3] Imran, A., Uddin Ahmed, M., Hussain, S. & Iqbal, J. (2022). Social engagement analysis for detection of fake news on twitter using machine learning. Vol 19.

[4] Li, J. & Lei, M. (2022). A brief survey for fake news detection via deep learning models. In procedia computer science vol 214 pp 1339–44. Elsevier b.v.

[5] Limon, F. A. & Jahan, N. (2021). Fake news detection using deep learning. http://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/7430

[6] Pennycook, G. & Rand, D. G. (2021). The psychology of fake news. Trends cogn sci 25, 388–402.

[7] George, J., Gerhart, N. & Torres, R. (2021). Uncovering the truth about fake news: a research model grounded in multi-disciplinary literature. Journal of management information systems 38, 1067–94.

[8] Hu, B., Mao, Z. & Zhang, Y. (2024). An overview of fake news detection: from a new perspective. Fundamental research. ISSN 2667-3258, https://doi.org/10.1016/j.fmre.2024.01.017.

[9] Burgers, N., Imaad, T. & Aladeen, H. (2023). Can machine learning algorithms really stop fake news in its tracks?

[10] Hanshal, O. A., Ucan, O. N. & Sanjalawe, Y. K. (2023). Hybrid deep learning model for automatic fake news detection. Applied nanoscience (switzerland) 13, 2957–67.

[11] John, A. & Journals, S. (2022). Fake news detection using n-gram analysis and machine learning algorithms.

[12] Jain, A., Shakya, A., Khatter, H. & Gupta, A. K. (2019). A smart system for fake news detection using machine learning. In ieee international conference on issues and challenges in intelligent computing techniques, icict 2019. Institute of electrical and electronics engineers inc.

[13] Melvern, A., Sibaroni, Y. & Prasetiyowati, S. S. (2023). Fake news detection: hybrid deep supervised learning approach. In 2023 international conference on data science and its applications, icodsa 2023 pp 414–9. Institute of electrical and electronics engineers inc.

[14] Güler, G. & Gündüz, S. (2023). Deep learning based fake news detection on social media. International journal of information security science 12, 1–21.

[15] Asta, R. S. & Budi Setiawan, E. (2023). Fake news (hoax) detection on social media using convolutional neural network (cnn) and recurrent neural network (rnn) methods. In international conference on ict convergence vol 2023-August pp 511–6, Ieee computer society.

[16] Aslam, N., Ullah Khan, I., Alotaibi, F. S., Aldaej, L. A. & Aldubaikil, A. K. (2021). Fake detect: a deep learning ensemble model for fake news detection. Complexity 2021.

[17] Tin, P. T. (2018). A study on deep learning for fake news detection. JAIST: Japan Advanced Institute of Science and Technology, 1-49.

[18] Verma, P. K., Agrawal, P., Amorim, I. & Prodan, R. (2021). Welfake: word embedding over linguistic features for fake news detection. Ieee trans comput soc syst 8 881–93.

[19] Alyoubi, S., Kalkatawi, M. & Abukhodair, F. (2023). The detection of fake news in arabic tweets using deep learning. Applied sciences (switzerland) 13.

[20] Hailemichael, E. N. (2021). Fake news detection for amharic language using deep learning adama, ethiopia.

[21] Thiyagarajan, V. S. (2024). Identifying fake news on isot data using stemming method with a subdomain of ai algorithms. Migration letters, 21 (S6), 775-787.

[22] Mjaaland, H. L., Vinay, R., Setty, J. & Anand, A. (2020). Detecting fake news and rumors in twitter using deep neural networks vinay jayarama setty.

[23] Ahmad, I., Yousaf, M., Yousaf, S. & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. Complexity 2020.

[24] Abdullah-All-Tanvir, Mahir, E. M., Akhter, S. & Huq, M. R. (2019). 2019 7th international conference on smart computing & communications (icscc).

[25] Khanam, Z., Alwasel, B. N., Sirafi, H. & Rashid, M. (2021). Fake news detection using machine learning approaches. Iop conf ser mater sci eng 1099 012040.

[26] Choudhary, A. & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. Expert syst appl 169.

[27] Reddy, H., Raj, N., Gala, M. & Basava, A. (2020). Text-mining-based fake news detection using ensemble methods. International journal of automation and computing 17, 210–21.

[28] ANON. Www.uvic.ca/engineering/ece/isot/assets/docs/isot_fake_news_dataset_readme.pdf.

[29] Ali, A. M., Ghaleb, F. A., Al-Rimy, B. A. S., Alsolami, F. J. & Khan, A. I. (2022). Deep ensemble fake news detection model using sequential deep learning technique. Sensors 22.

[30] Samadi, M., Mousavian, M. & Momtazi, S. (2021). Deep contextualized text representation and learning for fake news detection. Inf process manag 58.

[31] Goldani, M. H., Momtazi, S. & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. Appl soft comput 101.

[32] Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R. & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. Future generation computer systems 117, 47–58.

[33] Goldani, M. H., Safabakhsh, R. & Momtazi, S. (2021). Convolutional neural network with margin loss for fake news detection. Inf process manag 58.

[34] Gifu, D. (2023). An intelligent system for detecting fake news. In procedia computer science vol 221, pp 1058–65. Elsevier b.v.