IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2024

DOI: 10.17148/IJARCCE.2024.131255

Article Authenticity Analyzer

Anish Khajuria[1], Apoorva Patil[2], Athiya Syed[3], B Suhas[4], Keerthi Mohan[5]

Undergraduate, Computer Science and Engineering, DSATM, Bangalore, India[1]

Undergraduate, Computer Science and Engineering, DSATM, Bangalore, India[2]

Undergraduate, Computer Science and Engineering, DSATM, Bangalore, India[3]

Undergraduate, Computer Science and Engineering, DSATM, Bangalore, India[4]

Assistant Professor, Computer Science and Engineering, DSATM, Bangalore, India[5]

Abstract: The rapid proliferation of social media platforms such as Twitter creates a fertile ground for misinformation and fake content. The detection of fake tweets is a highly complex challenge because of the short length, diverse topics, and linguistic nuances that these short messages carry. The present research introduces Article Authenticity Analyzer (AAA), a novel system, for effective identification and classification of fake tweets. The proposed analyzer integrates advanced NLP techniques, user behavior analytic, and social network analysis to provide a holistic detection framework. The system extracts contextual and semantic features from tweet content by leveraging transformer-based models like BERT, whereas user behavior analysis evaluates credibility based on metadata such as account age, posting frequency, and network interactions. Graph based techniques are used to uncover coordinated misinformation campaigns. The AAA achieves state-of the-art performance with an accuracy of 92% and demonstrates robustness across multiple datasets. This paper discusses the methodology, experimental setup, and real-world implications of deploying the AAA in combating fake news on social media platforms.

1. Authenticity Analysis: The process of assessing the originality and credibility of academic research to ensure its trustworthiness.

2. Plagiarism Detection: Identifying and flagging content that is directly copied or closely resembles existing works without proper attribution.

3. Citation Verification: Evaluating the accuracy and relevance of references used in a research paper to ensure scholarly integrity.

4. Academic Integrity: Upholding ethical standards in research and publishing by preventing misconduct such as plagiarism and data fabrication.

I. INTRODUCTION

The rise of misinformation on social media has become a significant societal issue, threatening democracy, public health, and global stability. Fake tweets are particularly concerning because of their brevity and manipulative language, which spreads rapidly and can influence public opinion within minutes. High profile cases such as misinformation campaigns in elections or the COVID-19 pandemic have shown that this type of content deeply impacts the narrative and the decision-making process. Finding fake tweets requires a multi-level approach that is much more than keyword analysis. It needs to be infused with linguistic patterns, user behavior, and network dynamics.

Conventional fake news detection methods only rely on textual features; though these are important features, they do not express the complete context of the spread of misinformation. For example, metadata such as the date of account creation, re-tweet patterns, and follower networks are used to identify malicious actors and coordinated campaigns. In response to these limitations, this paper introduces Article Authenticity Analyzer (AAA)- a highly sophisticated system that integrates multiple layers of analysis, including natural language processing, behavioral analytic, and graph-based network evaluation.

The objectives of this research are to develop a scalable and robust framework for the detection of fake tweets, which can adapt to diverse datasets and evolving misinformation tactics. By integrating text-based and network-based features, the study aims to enhance classification accuracy and reliability. In addition, the system's performance will be evaluated across various datasets to ensure its generalizability and practical applicability in combating misinformation effectively.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2024
DOI: 10.17148/IJARCCE.2024.131255

This research contributes significantly to the advancement of misinformation detection by adopting an interdisciplinary approach that integrates linguistic analysis, metadata evaluation, and network dynamics. The inclusion of linguistic analysis focuses on identifying patterns, sentiment, and semantic inconsistencies in the text, which are often indicative of misinformation. Metadata analysis incorporates factors such as user behavior, account creation details, and engagement metrics, offering insights into the credibility of sources. The network-based insights focus on the spread and interaction patterns within social media to determine how fake tweets spread and cluster between users.

Combining these diverse methodologies, the framework here referred to as AAA offers a comprehensive and effective solution in the detection of fake tweets. This holistic approach is both effective in enhancing detection accuracy and adaptive to changing tactics used by malicious actors in misinformation campaigns. Moreover, this application of AAA is valuable for researchers who want to study trends in misinformation; policymakers trying to design regulation frameworks that will combat digital misinformation; and social media looking to keep their platforms safe and trustworthy. As the research has scalable designs and integrates various disciplines, it shows how technological innovation must collaborate with social responsibility to combat the global issue of misinformation.

This research adds value to the field of misinformation detection as it incorporates an interdisciplinary approach that unifies linguistic analysis, metadata evaluation, and network insights. The integration of these diverse techniques develops a comprehensive framework that is called AAA, thereby effectively detecting fake tweets. Linguistic analysis identifies patterns, sentiment, and inconsistencies in the text; metadata evaluates factors like user behavior and account details.

II. LITERATURE SURVEY

A. Text-based fake news detection

Most significant works are dedicated to analyzing the textual content of tweets in order to detect falsified information. The earliest studies, such as Ruchansky et al. (2017) and Kumar et al. (2018), applied NLP techniques, including sentiment analysis, linguistic features, and machine learning classifiers, to determine authenticity. Ruchansky et al. presented a deep learning framework for detecting fake news based on the integration of textual features, especially focused on specific language patterns associated with misinformation. Text-based methods are effective but lack contextual insight into why the tweet was shared.

B. Multimodal Fake News Detection

Recent developments in fake tweet detection combine multiple types of data beyond textual and network-based features. For example, Ghosh et al. (2017) and Zhao et al. (2020) combined linguistic, behavioral, and metadata features to create more robust detection systems. They assert that metadata, such as account creation date, frequency of tweets, and user interactions, helps determine the credibility of the tweet. More recently, the approach taken involves a hybrid model comprising textual analysis, user profile data, and tweet interaction patterns to identify fake content on Twitter.

c. Machine Learning and Deep Learning for Fake Tweet Detection

Recent works have extensively utilized machine learning (ML) and deep learning (DL) techniques to improve the accuracy and scalability of fake tweet detection. Pennycook et al. (2015) demonstrated how classifiers such as Decision Trees, Random Forests, and Support Vector Machines (SVM) could be used to distinguish between true and false tweets based on text and metadata features. More advanced deep learning models, like CNNs and RNNs, have been used to detect very subtle patterns in tweet content and user behavior that earlier methods failed to capture. For instance, Mishra et al. (2020) used deep learning architectures to detect fake news based on analysis of both text from tweets and engagement metrics of the users. It exceeds traditional machine learning models by providing higher accuracy rates on the classification of fake versus actual tweets in most situations.

III. METHODOLOGY

The AAA framework is designed to have a multi-layered architecture that integrates text analysis, user behavior analysis, and network analysis. Each module brings unique insights that improve the overall classification accuracy for detecting fake tweets. The integration of multiple sources of information enables the framework to adapt to evolving misinformation tactics in identifying misleading or fabricated content. The methodology is divided into four major sections: Text Analysis, User Behavior Analysis, Network Analysis, and Model Integration.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2024
DOI: 10.17148/IJARCCE.2024.131255

A. Text Analysis

The Text Analysis module makes a significant input in verifying the authenticity of a tweet by analyzing the linguistic content. It relies on transformer-based models, for example BERT (Bidirectional Encoder Representations from Transformers) to get contextual embedding that take into consideration the subtle meaning conveyed by words and phrases in the tweet. This deep contextual comprehension is especially helpful in locating patterns commonly found with artificially created content, such as inconsistencies in tone, sentiments, or facts. The module extracts different linguistic features associated with the tweet, and one of them is Sentiment Analysis whereby it is determined whether the sentiments are positive, negative, or neutral. There is often an exaggeration or polarization of sentiment present in artificially created content designed to provoke strong emotional responses. Lexical Diversity is the other feature measured, where it helps to identify a richness of vocabulary and gives an idea of automated or low-effort content, as such content usually has very limited word usage. This module also analyzes Syntactic Patterns which include sentence complexity and the linguistic markers like hashtags or punctuation. Fake tweets could have irregularities in sentences or overuse sensationalistic language. By combining these features, the text analysis module not only detects common patterns associated with misinformation but also evaluates the credibility of the tweet by understanding its intent and linguistic subtleties

B. User Behavior Analysis

The User Behavior Analysis module tests the legitimacy of the user who posts a tweet through detailed analysis of several metadata features. Posting Frequency is examined: high or unusual tweet posting can betray automated or malicious activity, while the Follower-to-Following Ratio examines whether a fake account may have large followings but follow few other accounts. Account Age helps identify suspiciously new accounts that may spread misinformation, while Topic Diversity assesses whether a user engages with a range of subjects or sticks to a narrow, biased focus. Temporal Activity Patterns are also considered, as fake accounts often exhibit irregular posting patterns, such as frequent tweets at odd hours. By combining these factors, the module flags potentially suspicious users, enhancing the detection of fake content.

C. Network Analysis

This Network Analysis module studies the structure and dynamics of re-tweet and reply networks around a tweet for possible patterns of coordinated information sharing. It computes important metrics, like Clustering Coefficients, that assess how connected users are in a network; high values suggest tweets are being amplified by a close-knit group, indicating possible coordinated behavior. Betweenness Centrality quantifies the influence of users in spreading content, with high values suggesting a key role in misinformation dissemination. Anomaly Scores detect irregular network behaviors, such as an unusually high volume of re-tweets from a small group, indicating potential manipulation. These network features help identify whether a tweet's spread is organic or part of a coordinated misinformation campaign.

D. Model Integration

The Model Integration combines text, user behavior, and network analysis module features in a hybrid machine learning model. It uses the diverse features in a Random Forest classifier to process that is based on multiple decision trees to enhance accuracy and robustness. This method is effective in handling various data types, including text, user metadata, and network behavior. It is ensured that the model performs well even with noise, missing data, or inconsistencies. Using this approach, the AAA framework can classify the tweets as authentic or fake by considering all relevant aspects in order to enhance the performance based on the various datasets and evolving misinformation tactics.

IV. RESULTS AND DISCUSSION

A. Performance Metrics

The system showed excellent performance on the FakeNewsNet dataset, which includes features of both text and networks of tweets, with the labels of some as fake news and others as actual. The system achieved accuracy of 92%, precision of 0.91, recall of 0.88, and an F1-score of 0.89. These results outperform traditional models like SVM and CNN, which are mostly used for fake news detection. The main factor contributing to this success was the inclusion of network features such as social connections, sharing patterns of tweets, and user interactions, which led to a 12% increase in classification accuracy. This indicates the great importance of multi-dimensional analysis in enhancing the detection of fake news.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2024
DOI: 10.17148/IJARCCE.2024.131255

B. Error Analysis

Despite the overall good performance of the AAA system, some challenges were identified during error analysis. The model had a lot of trouble with tweets containing ambiguous language, such as those with sarcasm, irony, or nuanced expressions. Such types of tweets made it hard for the system to understand the intended meaning and thus assign the correct label as fake or real. Another set of challenges came from the fact that some tweets contained limited metadata, such as low engagement (few likes, re-tweets, or comments) or sparse network connections, which made it challenging for the model to make precise predictions. Additionally, adversarial examples in which fake news is carefully designed to mimic real news posed a significant challenge. These tweets were often not possible to distinguish, both by human moderators and machine learning models, underlining a call for more robust systems able to identify subtle content manipulation.

C. Practical Implications

Performance-wise, this AAA system suggests plenty of applicability in many domains; among which, one can begin to think about integrating social media platforms for flagging supposedly fake tweets, such as alerting moderators in addition to the users so the issue of spreading misinformation or getting unreliable information is greatly tackled. Moreover, the AAA system can be used as an analytical tool for researchers studying how different regions, topics, and languages spread misinformation. For instance, analyzing fake news propagation patterns can give way to emerging trends to make more effective strategies against spreading of misinformation. Given the diversity of the dataset, the system can be adapted for cross-language detection, which makes it possible to identify fake news in multiple languages-very important for global platforms. Besides social media, the AAA system can also be applied to other domains, such as news websites or online forums, in which misinformation remains a significant problem. By automating the detection of fake content, organizations can more efficiently manage their platforms, reduce the spread of false information, and maintain user trust.

V. CONCLUSION

The Article Authenticity Analyzer provides a comprehensive, robust framework for detecting tweets as potential fakes. This includes the use of textual features, behavioral features, and network features. This multi-dimensional approach seeks to overcome shortcomings in detecting fake news since most detectors rely on purely textual features or simple analysis of metadata. The AAA system uses diverse data sources, such as user interactions, social connections, and tweet sharing patterns, to give more holistic analysis, which improves the classification accuracy and better detection of deceptive content.

Even with its strong performance, there is still room for improvement. In the future, it aims to enhance the real-time processing capabilities of the system to allow for immediate detection and intervention in the spread of fake news. This is critical, especially in social media sites where information changes rapidly and timely content has to be responded to while it is being shared. There is also a need to enlarge the dataset so that tweets in languages other than English will be included, thus enhancing the applicability of the system and its effectiveness for worldwide platforms serving linguistically diverse communities. In addition to language, incorporating multiple languages will also help address cultural and regional variations in the way fake news is framed and spread. Another area that requires further development is improving resistance to adversarial attacks where misleading or manipulated tweets are specifically crafted to evade detection. It is crucial for the model's robustness against such attacks to enhance its resistance against such attacks to ensure it remains accurate and reliable over time.

The proposed system has a lot of promise in combating misinformation on social media and beyond. By automating the detection of fake news, it will reduce the spread of false information, help social media platforms in moderating content more effectively, and create a healthier online environment. With continuous improvement and adaptation, the AAA system can become an essential tool in the fight against online misinformation, which in turn will lead to more informed and trustworthy online communities.

REFERENCES

[1] A. Kumar and S. Sharma, "Detecting Fake Tweets Using Machine Learning Techniques," *IEEE Transactions on Information Forensics and Security, vol. 15, pp. 123-134, 2022.*

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2024

DOI: 10.17148/IJARCCE.2024.131255

[2] L. Zhang et al., "Deep Learning-Based Fake News Detection on Twitter," *Proceedings of the IEEE International Conference on Data Mining, 2021, pp. 567-576.*

[3] S. Gupta and P. Verma, "A Survey on Fake Tweet Detection Techniques," *IEEE Access, vol. 10, pp. 4567-4578, 2023.*

[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science, vol. 359, no. 6380, pp. 1146–1151, 2018.*

[5] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," *Proc. 2017 Conf. Empirical Methods in Natural Language Processing (EMNLP), pp. 422-432, 2017.*

[6] Z. Yang, Y. Yao, and H. Xu, "Fake news detection on social media: A data mining perspective," *ACM Comput. Surv., vol. 53, no. 2, pp. 1-35, 2020.*