WELFAKE: WORD EMBEDDING OVER LINGUISTIC FEATURES FOR FAKE NEWS DETECTION

G KRISHNA LAVA KUMAR[1], U VAMSHI[2], U SAI KUMAR[3], G UDAY KIRAN[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT- Social media has gained immense popularity as a platform for disseminating real-time news globally, driven by the ease and speed of information sharing. However, this rapid proliferation of content has a significant downside—fake news. People often share news without verifying its authenticity, which has led to widespread misinformation. To address this pressing issue, this article presents a two-phase benchmark model, WELFake, designed to detect fake news using machine learning classification with word embedding (WE) and linguistic features. In the first phase, the model preprocesses the dataset and verifies the authenticity of news content by leveraging linguistic features. The second phase combines these linguistic features with word embeddings and applies a voting classification method to improve the accuracy of predictions. A novel dataset called WELFake, consisting of approximately 72,000 articles from diverse sources, was created to evaluate the performance of the model. The experimental results demonstrate that the WELFake model achieves an impressive accuracy of 96.73% in classifying news as real or fake. This performance exceeds the accuracy of the Bidirectional Encoder Representations from Transformers (BERT) model by 1.31% and the Convolutional Neural Network (CNN) model by 4.25%. Furthermore, WELFake outperforms existing predictive-based approaches that use Word2Vec for word embeddings by up to 1.73%. These findings highlight the model's effectiveness in improving the accuracy of fake news detection and emphasize its potential to mitigate the spread of misinformation in the digital space. By focusing on linguistic features and a novel voting classification approach, WELFake offers a promising solution for more reliable news verification.

INDEX TERMS- Fake News Detection, Machine Learning, Word Embedding (WE), Linguistic Features, Classification, WELFake Model, News Verification, Data Preprocessing, Voting Classification, Bidirectional Encoder Representations from Transformers (BERT).

I. INTRODUCTION

Social media has become a dominant platform for global communication, attracting billions of users across different demographics. However, its ease of information dissemination also contributes to the rapid spread of fake news, which is intentionally false content propagated for manipulation, political gain, or financial profit. Fake news is harmful as it influences public opinion, spreads misinformation, and leads to negative societal consequences like poor decision-making and social unrest. Despite various websites like PolitiFact and Snopes aimed at verifying news, these platforms cannot respond promptly to new fake news events, highlighting the need for more efficient automated solutions.

Researchers have focused on using machine learning models to detect fake news, yet challenges persist in identifying the most effective linguistic features, selecting the best word embedding techniques, and choosing the optimal classification method for fake news detection.

In response to these challenges, the article proposes a novel fake news detection model called WELFake, which uses text data in three main stages. It first employs linguistic feature sets (LFS) for prediction, then applies word embedding (WE) over these features to improve detection, and finally conducts comparative analysis with state-of-the-art methods like CNN and BERT. The authors also created a larger WELFake dataset by merging four popular news datasets, which contains 72,134 articles (35,028 real and 37,106 fake news). The model achieved an accuracy of 96.73% in classifying real and fake news, outperforming CNN by 4.25% and BERT by 1.31%. The study demonstrates that combining LFS and WE techniques enhances the detection of fake news, and the model's generalization and effectiveness were validated through adversarial testing, showing its potential for real-world applications.

II.LITERATURE SURVEY

A) W. Jiang, J. Wu, F. Li, G. Wang, and H. Zheng, "Trust evaluation in online social networks using generalized network flow," IEEE Trans. Comput., vol. 65, no. 3, pp. 952–963, Mar. 2016.

In this paper, the authors propose a trust evaluation method for online social networks (OSNs) based on generalized network flow, aiming to address the challenge of trust management in these networks. Trust evaluation is critical in OSNs due to the large-scale interactions among users, where trust influences decision-making, reputation, and user behavior. The paper introduces a new framework that leverages generalized network flow to evaluate trust relationships more effectively by considering the complex interactions within the network. The proposed method takes into account both direct and indirect trust relationships, incorporating network flow concepts to model trust propagation in a more dynamic and scalable manner. The authors highlight the need for accurate trust evaluation to mitigate issues like fake identities and malicious behaviors in OSNs. Through experiments on real-world datasets, the method demonstrates its efficiency and effectiveness in comparison to existing trust evaluation techniques. The results show that the generalized network flow approach provides more reliable and accurate trust evaluations, contributing to the development of more secure and trustworthy online social platforms. This work has significant implications for improving trust management in OSNs and related applications, such as recommender systems and collaborative platforms.

B) M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino, "Credibility in online social networks: A survey," IEEE Access, vol. 7, pp. 2828–2855, 2019.

This paper provides a comprehensive survey on credibility in online social networks (OSNs), addressing the growing concern about the reliability of information shared in these platforms. As OSNs become essential for information dissemination, the spread of false, misleading, or biased content has become a significant issue. The authors discuss various approaches to assess and evaluate credibility in OSNs, including trust-based, content-based, and user-based methods. They explore existing algorithms and techniques that aim to identify credible sources, detect misinformation, and manage reputation within social networks. The paper also reviews credibility challenges specific to OSNs, such as the influence of bots, fake accounts, and malicious users. Moreover, it highlights the impact of credibility on decision-making, user behavior, and overall network performance. The authors emphasize the need for effective tools and systems to address credibility issues and improve the quality of information shared in OSNs. Finally, the paper presents a critical analysis of the limitations and open challenges in the field, suggesting future research directions to enhance credibility management in online social platforms.

C) S. Ranganath, S. Wang, X. Hu, J. Tang, and H. Liu, "Facilitating time critical information seeking in social media," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2197–2209, Oct. 2017

This paper addresses the challenge of facilitating time-critical information seeking in social media, where the rapid spread of information can significantly impact real-time decision-making. The authors propose a framework to help users efficiently find relevant information during time-sensitive situations, such as emergencies or breaking news events. The framework includes strategies for ranking and filtering social media content, ensuring that users can quickly access trustworthy and pertinent information. The paper examines the key factors influencing information search in social media, including content relevance, user engagement, and content credibility. The authors also highlight the importance of context-awareness and social network structures in improving information retrieval. Furthermore, they discuss the trade-offs between accuracy and timeliness when seeking time-critical information. The proposed approach incorporates machine learning techniques to enhance the relevance and effectiveness of the search process. The results demonstrate the potential for improving information access during high-pressure situations, emphasizing the need for better tools in dynamic online environments.

IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

In this work, the authors introduce the WELFake model for fake news detection, aiming to improve classification accuracy by combining linguistic features and word embedding (WE) techniques. They created a comprehensive dataset, WELFake, by merging four open-source datasets (Kaggle, McIntire, Reuters, and BuzzFeed), reducing individual dataset limitations. Over 80 linguistic features were analyzed, and 20 significant features were selected to optimize computational efficiency and classifier accuracy. Two WE methods, TF-IDF and CV, were applied across six machine learning models (SVM, KNN, Naive Bayes, Decision Trees, Bagging, and AdaBoost), with CV outperforming TF-IDF when coupled with the SVM model. The selected linguistic features were categorized into writing pattern, readability index, psycho-linguistics, and quantity. The features were distributed into three balanced sets for voting classification, which incorporated results from each WE-enabled LFS dataset. The final classification was determined using a second-level voting classifier, which showed that the WELFake model achieved an impressive 96.73% accuracy. This result outperformed BERT by 1.31% and CNN by 4.25%, and improved performance by up to 10% on certain datasets. The SVM model was found to be the most accurate, and the frequency-based model, focusing on writing patterns, outperformed predictive models such as Word2Vec by 1.73%. Future work will explore integrating knowledge graphs and user credibility for enhanced fake news verification.

REFERENCES

[1] W. Jiang, J. Wu, F. Li, G. Wang, and H. Zheng, "Trust evaluation in online social networks using generalized network flow," IEEE Trans. Comput., vol. 65, no. 3, pp. 952–963, Mar. 2016.

[2] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino, "Credibility in online social networks: A survey," IEEE Access, vol. 7, pp. 2828–2855, 2019.

[3] S. Ranganath, S. Wang, X. Hu, J. Tang, and H. Liu, "Facilitating time critical information seeking in social media," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2197–2209, Oct. 2017.

[4] Z. Zhang, R. Sun, X. Wang, and C. Zhao, "A situational analytic method for user behavior pattern in multimedia social networks," IEEE Trans. Big Data, vol. 5, no. 4, pp. 520–528, Dec. 2019.

[5] M. Schudson and B. Zelizer, "Fake news in context," in Understanding and Addressing the Disinformation Ecosystem. Philadelphia, PA, USA: Annenberg School for Communication, Apr. 2017, pp. 1–4.

[6] S. Zaryan, "Truth and trust: How audiences are making sense of fake news," M.S. thesis, Media Commun. Studies, Lund Univ. Publications Student Papers, Stockholm, Sweden, Jun. 2017. [Online]. Available: https://lup.lub.lu.se/student-papers/search/publication/8906886

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

[8] R. Sequeira, A. Gayen, N. Ganguly, S. K. Dandapat, and J. Chandra, "A large-scale study of the Twitter follower network to characterize the spread of prescription drug abuse tweets," IEEE Trans. Comput. Social Syst., vol. 6, no. 6, pp. 1232–1244, Dec. 2019.

[9] Politifact News Dataset. Accessed: Mar. 31, 2020. [Online]. Available: http://www.politifact.com/

[10] The Washingtonpost Fact Checker. Accessed: Mar. 31, 2020. [Online]. Available: https://www.washingtonpost.com/news/fact-checker

[11] Fact Check. Accessed: Mar. 31, 2020. [Online]. Available: https://www. factcheck.org/

[12] Snopes. Accessed: Mar. 31, 2020. [Online]. Available: https://www. snopes.com/

[13] Truthorfiction. Accessed: Mar. 31, 2020. [Online]. Available: https:// www.truthorfiction.com/

[14] Fullfact. Accessed: Mar. 31, 2020. [Online]. Available: https://fullfact. org/

[15] Hoax Slayer. Accessed: Mar. 31, 2020. [Online]. Available: http://hoaxslayer.com/