

Fake news detection on social media using Natural Language Processing : A Systematic Review

Dharmendra Singh, Shiva Soni, Kumkum Soni, Hanshraj Singh

Department of Computer Science AKS University, Satna, M.P
Email: dharamsingh62607@gmail.com

Abstract— The rapid growth of social media platforms has significantly increased the spread of misinformation, posing serious challenges to information credibility and public trust. Due to the massive volume of online content, manual verification is impractical, making automated fake news detection essential. This review paper presents a comprehensive analysis of Natural Language Processing (NLP) techniques used for fake news detection. It examines both traditional machine learning models, such as TF-IDF with Logistic Regression and Passive Aggressive Classifier, as well as advanced deep learning approaches including BiLSTM and BERT. The findings indicate that transformer-based models, particularly BERT, achieve superior performance due to their contextual understanding, while lightweight models offer efficiency for real-time applications. Despite these advancements, challenges such as model interpretability, domain adaptation, and multilingual detection remain open for future research.

Index Terms—Fake News Detection, Natural Language Processing (NLP), Machine Learning, Social Media Analysis, Passive Aggressive Classifier, TF-IDF, ISOT Dataset

I. INTRODUCTION

In recent years, social media platforms have become a major source of information [1], enabling rapid sharing of content without traditional editorial control. While this improves accessibility, it also increases the spread of misinformation.

Fake news refers to misleading or false information presented as authentic [1], [5], and it has become a serious concern due to its impact on public trust, democratic processes, and social stability. Its rapid spread is further amplified by social media algorithms and echo chambers.

The large volume and speed of online data make manual verification ineffective, creating a need for automated detection systems. To address this, researchers use Artificial Intelligence (AI) and Machine Learning (ML), with Natural Language Processing (NLP) playing a key role in analyzing textual data and identifying deceptive patterns.

II. LITERATURE REVIEW

Fake news detection has evolved significantly with advancements in Natural Language Processing techniques. Early approaches relied on manual verification and rule-based systems [3], which lacked scalability. Subsequently, machine learning models utilizing feature extraction methods

such as Bag-of-Words and TF-IDF, combined with classifiers like Naïve Bayes and Support Vector Machines, provided moderate performance [8]. However, these models were limited in capturing semantic and contextual relationships.

Recent studies have shifted toward deep learning and transformer-based models [6], [10], including CNN, RNN, BiLSTM, and BERT, which offer improved accuracy by understanding complex linguistic patterns. Additionally, ensemble and hybrid models have further enhanced detection performance [20]. Despite these improvements, challenges such as data quality, model interpretability, and cross-domain generalization remain critical issues.

TABLE I. : LITERATURE REVIEW SUMMARY

Author	Year	Technique	Dataset used	Key Finding
Mayank et al.	2021	NLP + GNN	Politifact, GossipCop	Better accuracy using knowledge graphs
Sutradhar et al.	2022	ML + NLP	Custom dataset	ML works but less accurate than DL
Shaik et al.	2023	NLP + PA Classifier	Kaggle Fake News Dataset	Fast and efficient results
Narayanan et al.	2023	Transformer (BERT)	NELA-GT	High contextual accuracy
Almandouh et al.	2024	Ensemble DL	Twitter / Social Media datasets	Higher accuracy than single models
Roumeliotis et al.	2025	CNN, BERT, LLMs	Benchmark datasets	BERT/LLMs perform best

III. METHODOLOGY

The proposed methodology follows a structured pipeline for fake news detection using Natural Language Processing (NLP). It begins with collecting benchmark datasets such as FakeNewsNet, LIAR, and Kaggle Fake News datasets, which contain labeled examples of real and fake news articles.

The data undergoes preprocessing, where noise such as punctuation, URLs, and special characters is removed. Text

normalization includes lowercase conversion, tokenization, and lemmatization. Missing values are handled, and the dataset is split into training (80%) and testing (20%) sets.

Next, feature extraction methods are applied to transform textual data into numerical representations. These include TF-IDF (Term Frequency–Inverse Document Frequency) for identifying important words, word embeddings (Word2Vec and GloVe) for capturing semantic relationships, and contextual embeddings such as BERT for deep contextual understanding [2], [7].

The core of the system lies in the modeling methods, where multiple algorithms are used and compared. These include:

- BERT (Transformer-based model) – deep contextual language understanding
- Bidirectional LSTM (BiLSTM) – capturing sequential patterns
- Logistic Regression – baseline machine learning model
- Support Vector Machine (SVM) – effective text classification
- Random Forest – improving robustness through ensemble learning

Finally, the models are evaluated using standard performance metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC. This methodology ensures a reliable, comparative, and data-driven approach.

IV. RESULT AND DISCUSSION

BERT achieved the highest performance, followed by BiLSTM, while Logistic Regression demonstrated comparatively lower accuracy. These results indicate that deep learning models with strong contextual understanding outperform traditional machine learning approaches [5], [6]. However, challenges such as sarcasm detection and handling complex linguistic structures still remain.

TABLE II. : MODEL PERFORMANCE COMPARISON

MODEL	ACCURACY	PRECISION
LOGISTIC REGRESSION (TF-IDF)	0.83	0.82
BiLSTM (WORD2VEC)	0.88	0.87
BERT FINE-TUNED	0.98	0.93

V. CONCLUSION

This review highlights the effectiveness of Natural Language Processing techniques in detecting fake news on social media platforms. Among various approaches, transformer-based models such as BERT demonstrate superior performance due to their ability to capture contextual and semantic information.

Traditional machine learning models, although less accurate, remain useful for lightweight and real-time applications. The

study emphasizes that combining syntactic and semantic features enhances detection capabilities. Future research should focus on multimodal analysis, improving model interpretability, and developing efficient models for real-time and multilingual environments.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Dr. Virendra Tiwari for his valuable guidance, constant support, and encouragement throughout the preparation of this review paper on “Fake News Detection on Social Media using Natural Language Processing (NLP).”

His insightful suggestions and expert knowledge greatly helped us in understanding the topic and organizing our work effectively. We are especially thankful for his continuous motivation and for clarifying our doubts during the review process.

We also acknowledge the contributions of various researchers whose work has been referenced in this paper, as their studies provided a strong foundation for our analysis. Finally, we thank all group members and supporters who contributed directly or indirectly.

REFERENCE

- [1] B. Dhiman, “The rise and impact of misinformation and fake news on digital youth: A critical review,” SSRN Electronic Journal, 2023.
- [2] D. De, B. Bandyopadhyay, B. Gain, and A. Ekbal, “A transformer-based approach to multilingual fake news detection in low-resource languages,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 21, no. 1, 2022.
- [3] H. Jand, “Natural language processing with speech recognition,” *Int. J. Adv. Res. Trends Eng. Technol.*, 2018.
- [4] P. Bhardwaj, P. Khanna, S. Kumar, and Pragya, “Generative model for NLP applications based on component extraction,” *Procedia Comput. Sci.*, 2020.
- [5] M. Tsai, “Stylometric fake news detection based on natural language processing using named entity recognition: In-domain and cross-domain analysis,” *Electronics*, vol. 12, no. 17, 2023.
- [6] R. Mohawesh, X. Liu, H. M. Arini, Y. Wu, and H. Yin, “Semantic graph-based topic modelling framework for multilingual fake news detection,” *AI Open*, vol. 4, 2023.
- [7] Y. Zhang et al., “Stance-level sarcasm detection with BERT and stance-centered graph attention networks,” *ACM Trans. Internet Technol.*, vol. 23, no. 2, 2023.
- [8] K. Sharifani, M. Amini, Y. Akbari, and J. A. Godarzi, “Operating machine learning across natural language processing techniques for improvement of fabricated news model,” *Int. J. Sci. Inf. Syst. Res.*, vol. 12, no. 9, 2022.
- [9] M. Q. Alnabhan, P. Branco, and M. Alnabhan, “Fake news detection using deep learning: A systematic literature review,” 2024.
- [10] J. Alghamdi, Y. Lin, and S. Luo, “Fake news detection in low-resource languages: A novel hybrid summarization approach,” *Knowl-Based Syst.*, vol. 296, 2024.
- [11] M. J. G. Fagundes, N. T. Roman, and L. A. Digiampietri, “The use of syntactic information in fake news detection: A

- systematic review,” SBC Rev. Comput. Sci., vol. 4, no. 1, pp. 1–10, 2024.
- [12] P. Krasadakis, E. Sakkopoulos, and V. S. Verykios, “A survey on challenges and advances in natural language processing,” MDPI, 2024.
- [13] P. Meesad, “Thai fake news detection based on information retrieval, natural language processing and machine learning,” SN Comput. Sci., vol. 2, no. 6, 2021.
- [14] K. Agras and B. Atay, “A novel transformer-based deep learning pipeline for multilingual fake news detection,” 2024.
- [15] R. F. Cekinel, P. Karagoz, and C. Coltekin, “Cross-lingual learning vs. low-resource fine-tuning: A case study with fact-checking in Turkish,” 2024.
- [16] S. Han, “Cross-lingual transfer learning for fake news detector in a low-resource language,” 2022.
- [17] M. Zovikoğlu and U. Çetin, “Detecting misinformation on social networks with natural language processing,” 2024.
- [18] A. Ziyaden et al., “Text data augmentation and pre-trained language model for enhancing text classification of low-resource languages,” PeerJ Comput. Sci., vol. 10, 2024.
- [19] P. Mookdarsanit and L. Mookdarsanit, “COVID-19 fake news detection in Thai social texts,” Bull. Electr. Eng. Informat., vol. 10, no. 2, 2021.
- [20] F. Farhangian, R. M. O. Cruz, and G. D. C. Cavalcanti, “Fake news detection: Taxonomy and comparative study,” Inf. Fusion, vol. 103, 2024.
- [21] B. Hu, Z. Mao, and Y. Zhang, “An overview of fake news detection: From a new perspective,” 2024.
- [22] J. A. Nasir and Z. U. Din, “Syntactic structured framework for resolving reflexive anaphora in Urdu discourse using multilingual NLP,” KSII Trans. Internet Inf. Syst., vol. 15, no. 4, 2021.
- [23] S. Ranathunga et al., “Neural machine translation for low-resource languages: A survey,” ACM Comput. Surv., vol. 55, no. 11, 2023.
- [24] B. Ramesh Reddy et al., “Optimising the detection of fake news in multilingual exposition using machine learning techniques,” Proc. ICICCS, 2023.