J. Mountain Res. P-ISSN: 0974-3030, E-ISSN: 2582-5011 DOI: https://doi.org/10.51220/jmr.v21-S2.15
Vol. 21(S2), (2026) 125-131

Fake News Observation Dummy Based On Machine Learning

Bhanu Partap* • **Deepak Saini** • **Nisha Sharma** • **Apoorva Singh**

Department of Computer Science and Engineering, Quantum University, Roorkee Dehradun Highway Mandawar, Uttarakhand 247167

*Corresponding Author's Email Id: Bhanu8909@gmail.com

Received: 26.02.2025; Revised: 25.10.2025; Accepted: 19.11.2025

Abstract: The dissemination of inaccurate information via online media is a serious danger to public discourse, societal cohesion, and democratic processes. In this notepaper, we present a novel approximate for notice fake news articles by leveraging linguistic patterns extracted from their textual content. Our technique ology involves the apply of advanced machine learning skill to analyze linguistic quality, such as verbal, syntactic, and semantic characteristics, in order that discern between genuine and fabricated news stories. We present a comprehensive examination of various linguistic indicators associated with fake news, including sensationalism, polarity, sentiment, and stylistic anomalies. Through extensive experimentation on diverse datasets sourced from reputable news outlets and known fake news sources, we evaluate the effectiveness of our proceed towards in correctly identifying deceptive content. Our findings re veal that linguistic patterns play a crucial role in distinguishing between actual and fake news articles, with certain quality, such as exaggerated language and emotional appeals, being more prevalent in deceptive content. Additionally, we investigate the robustness of our dummy across different domains, languages, and socio-political contexts, highlighting the generalizability of our proceed towards. Furthermore, we conduct a comparative examination with existing cutting-edge methods for observing bogus news to show how much better our language-based approach is. Our research contributes to the advancement of fake news observation technology by providing insights into the linguistic mechanisms underlying the dissemination of misinformation in online media environments. Moreover, our findings have practical implications for the design of automated fake news observation systems and the development of media literacy education programs aimed at fostering critical thinking skills among digital media consumers. By better understanding the linguistic cues indicative of fake news, we can enable people to successfully traverse the intricate web of information and reduce the harmful effects of misinformation on society. This paper bridges the gap between media studies, machine learning, and natural language processing, making it a major contribution to the interdisciplinary subject of computational journalism. Our goal is to create more powerful and efficient solutions to stop the spread of misleading information and preserve the accuracy of data in the digital age. through cooperation across disciplines.

Keywords: Fake news detection • Machine learning • Misinformation • Media literacy • Computational journalism • Sentiment analysis.

Introduction

The spread of social media and technology has changed how people receive and share information, but it has also led to the emergence of a worrying trend: fake news. In the current digital era, fake news—defined as purposefully false or deceptive content presented as actual news—has become a serious problem. The stability of society and democratic processes are seriously threatened by fake news due to its capacity to sway public opinion, foment strife, and erode trust in institutions. Preserving the integrity of information and guaranteeing that the public makes informed decisions require addressing the challenge of spotting fake news. While traditional fact-checking techniques have been employed to combat misinformation, they are often time-consuming, ineffective, and unable to keep up with the rapid spread of false material on the internet. In an attempt to get closer to automatic fake news identification, academics have turned to machine learning techniques in response to these challenges. Methods of machine learning can evaluate vast numbers of news items and more accurately

and efficiently identify fraudulent content by utilizing linguistic patterns and other textual quality indicators. With an emphasis on linguistic analysis, we present a unique method for machine learning-based fake news identification in this paper. Our objective is to offer a scalable and trustworthy technique for identifying fake news articles by utilizing linguistic cues extracted from their written content. Our goal is to positively influence ongoing efforts to prevent the spread of misleading material in online media spaces by utilizing cutting-edge methods in natural language processing and machine learning (Ahmed et al., 2021; Goyat et al., 2024; Pathan et al., 2023).

Literature Review

Several mechanisms for find fake news have recently been developed. The pertinent studies on spotting false information on social networking websites is included in this section. According to the available literature survey, Machine learning dummy's were often worn to identify fake news, then deep learning dummy's came into picture and nowadays transfer learning and pre-primed dummy's.

(Aldwairi et al. 2018) developed a browser-based tool to identify and eliminate misleading and clickbait content, emphasizing the necessity of multimodal observation since fake news involves altered text, audio, video, and images (Aldwairi et al., 2018). (Wang et al. 2018) suggested the Event Adversarial Neural Network (EANN) for multimodal fake news detection, which learns event-invariant features for real-time misinformation recognition on Twitter and Weibo (Wang et al., 2018). (Khan et al. 2019) analyzed different machine learning algorithms on large datasets and found that BERT-based models outperform others in fake news detection, even with small datasets (Khan et al., 2019). (Singhal et al. 2020, 2021) introduced multimodal models combining VGG-19 for images and BERT for text, demonstrating high accuracy on datasets like FakeNewsNet (Singhal et al., 2020). (Kaliyar et al. 2020) proposed Fake BERT, integrating CNN and BERT to enhance text understanding and outperform prior approaches (Kaliyar et al., 2020). (Zhou et al. 2020) presented SAFE, a model focusing on cross-modal similarity between text and images, confirming the superiority of deep learning over traditional machine learning methods (Zhou et al., 2020). (Ozbay and Alatas 2020, 2021) suggested hybrid AI-based approaches applying multiple supervised algorithms and swarm optimization (Ozbay and Alatas, 2020; 2021). (Sahoo et al. 2021) developed a deep learning-based Chrome extension for Facebook, where the LSTM model achieved 99.4% accuracy (Sahoo et al., 2021). (Collins et al. 2021) surveyed emerging detection strategies, emphasizing NLP, hybrid models, and graph-based techniques (Collins et al., 2021). Overall, the literature reveals an evolution toward advanced deep learning and transfer learning frameworks, emphasizing multimodal systems integrating textual, visual, and contextual cues using pre-trained architectures for more effective fake news detection (Ahmed et al., 2021; Kaur et al., 2020; Kumar et al., 2020).

Overall, literature trends reveal an evolution from machine learning to advanced deep learning and transfer learning frameworks. The focus has shifted toward multimodal systems integrating textual, visual, and contextual cues and leveraging pre-trained architectures such as BERT and CNN for more effective and generalizable fake news detection across diverse datasets and platforms.

Methodology

The classifying process is explained in this section. Using this dummy, a method is used to identify the phony articles. Supervised machine learning is used in this technique. to categorize the dataset. This classification challenge begins with the dataset offering phase. Following that, the preprocessing

proceeds smoothly, quality selection is completed, the dataset is tested and drilled into, and categorization is finally under control. Figure following reports the recommended system technique. The approach is built on handling several dataset tests using the previous section's reports on classification techniques, such as Naïve Bayes, SVM, Random Forest, most voting, and others.

Trials are performed sequentially on each algorithm and in combination with each other to attain the best possible accuracy and clarity. In order to wear the categorization, dummy the initial objective is to gather news observation details using a set of categorization approaches as a false news scanner. The dummy will then be worn as a detector for the bogus news data after being integrated into a Python application. Additionally, the necessary changes have been made to the Python code to produce An improved code. These techniques are all as accurate as they can be. were actual as a result of their averages being added up and equated. Figure 1 illustrates how different algorithms are applied to the dataset to detect false information. The precision of the findings collected is reviewed in order to complete the final result.

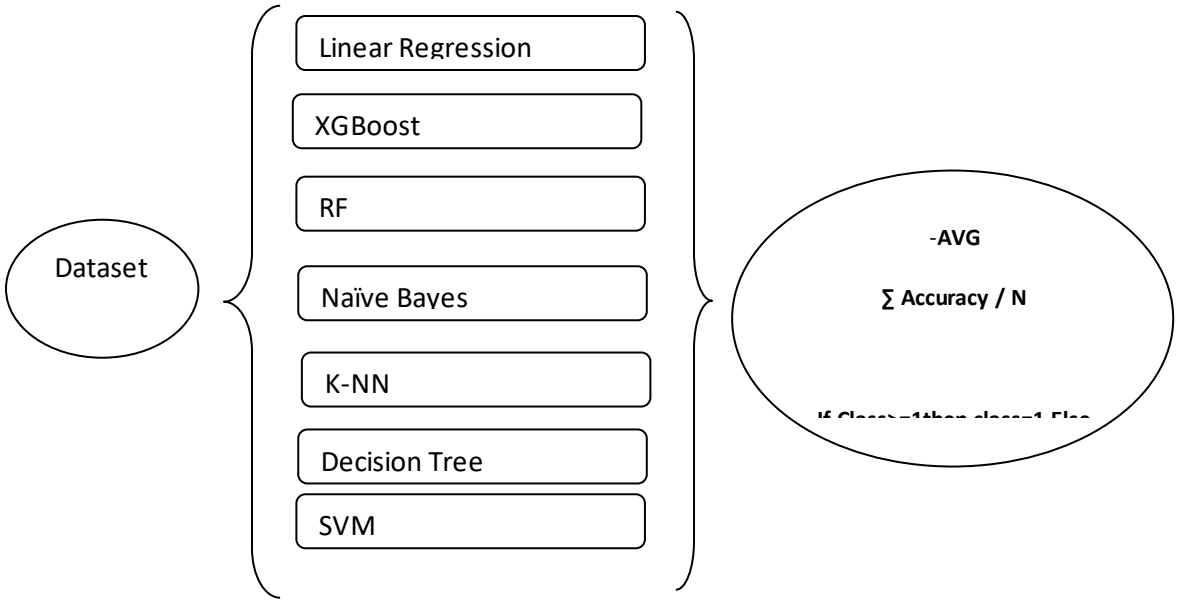

Figure 1. Various Methods

The following is where one can locate During the dummy blueprint process, political fake news: The group political news dataset is the initial phase (the Liar dataset serves as the dummy). Rough sound is eliminated as part of the pre-processing of the data. After that, POS and quality selection are carried out using the Natural Language Toolkit (NLTK). Implement the dataset break and create a recommend ranking dummy after using machine learning techniques (NBRF). After NLTK is applied, Figure 1 shows how the dataset is efficiently digested in the structure and how a message is generated for the primed component put in method. After applying the Random Forest and the system's N.B. response, a dummy with the response message is constructed. After a test dataset trial and a confirmed solution, the following stage is to monitor the clarity for acceptance. After the user selects hidden data, the dummy is real. Half of the information is utilized to construct the entire dataset. Half of the articles were real and half were bogus, giving the dummy a 50% reset accuracy. Once our dummy is complete, In our entire datasets, we use a random selection procedure to wear 80% of the data from the real and false datasets, with the remaining 20% being worn as a trial set. Before employing a classifier on text data, the data must be pre-processed. To

ensure that accomplish In addition to sound cleaning, Stanford Natural Language analysis (NLP) will be applied to Part of Speech (POS) analysis and word tokens. The output data must then be encoded as integers and floating-point numbers in order to be used as an input for the machine learning technique. This process will be sped up and traits will be eliminated by the research. The Python Scikit-Learn package, which includes useful utilities like the Tiff and Count vectorizers, will be used to tokenize and remove characteristics from text input. The data is displayed graphically along with a doubt grid. See figure 2.

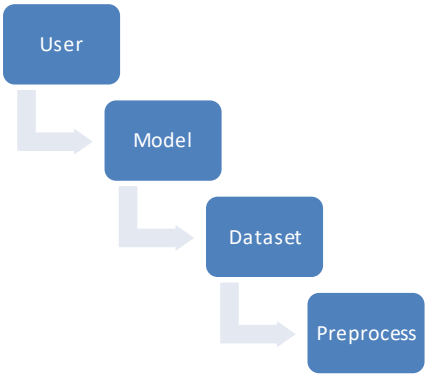

Figure 2. Model of Fake Detectors

The chosen datasets, the method, and the LIAR-PLUS Master used for data extraction and cleaning are covered in this section. This dataset has automatically eliminated page proof sentences written by journalists from the full-text verdict article report on PolitiFact.We donned the Truth values characteristic as shown in Figure 3. In order to gain four additional properties (nouns, verbs, prepositions, and sentences), we also applied part of speech to the declaration. Every piece of data has a class label. (0, 1, 2, 3) so that it may be worn when coaching the dummy. The following procedures have been used to assess how clear the news is.

1. Liar-datasets is a 12.8K digest
2. The texts are obtained from POLITIFACT.COM and manually tagged in various circumstances. Next, using Python, it is modified from the TSV pattern into CSV form.
3. The NLP NLTK and SAFAR v2 libraries must then be used to clean the sound. Commas, ids, dots, quotations, and the removal of the appendix are all suggested by the sound. The datasets are then converted into facts, services, and tokens using POS (Part of Speech).
4. Use a selected linguistic attribute to carry out characteristic removal. Examples include average word count and length, article length, number count, and quantity of adjective parts of speech.
5. To precisely extract Unigram and Bigram characters, use Python's sklearn's Vectorizer function. Make use of a characterisation TF-IDF n-gram quality is generated using a removal library.
6. Use Python Sklearn to separate the dataset into 70% for Prime and 30% for Test.
7. After using every strategy, create a dummy ipynb file for classification.
8. Make a doubt grid and test the test portion of the datasets for dummy clarity.
9. Examine the recall and f1-score., correctness, and clarity of real and bogus news.
10. Establish a connection with the user can wear to test hidden news.

Results

This project aims to protect political news data from being tagged as fake or trustworthy

news. It uses datasets called Liar-datasets, which are New Standard Datasets for Fake News Observation. We've conducted analysis on "Liar" datasets. The results of the dataset's evaluation using the six methodologies are displayed using the confusion grid. The six methods listed below were applied to the observation.:

- **XGBoost**: e**X**treme **G**radient **Boost**ing
- **Random Forests**: **RF**
- **Naive Bayes**: **NB**
- **K-Nearest Neighbour**: **KNN**
- **Decision Tree**: **DT**
- **Support Vector Machine**: **SVM**

Python code that utilizes the cognitive model automatically obtains the doubt grid when managing the approach code on the Anaconda platform. drilling module. The Confusion Grids for each technique are displayed in the graphic below:

Figure Results of each algorithm's accuracy Show how accurate different techniques can be in this figure. As can be observed, Random Forest and SVM both have approximately 73% accuracy, while XGBOOST has the greatest accuracy at over.

CONCLUSION

Characterization and disclosure are the two levels of analysis used in this study to investigate the detection of false news. In the first step, social media is used to emphasize the fundamental ideas and concepts of fake news. At the discovery stage, the efficacy of existing methods for detecting false news is assessed using multiple supervised learning algorithms. The paper uses fateful dummies that are unrelated to the other current dummies and dummies based on voice attributes to observe bogus news. The foundation of this strategy is text evaluation. With The classifier known as Naive Bayes, they were able to recognize fake news from various sources with an accuracy rate of 74%. The 85–91% accuracy prospect level of the Worn merging machine learning algorithm is reliant on unreliable elements makes use of Even though Naive Bayes was used to detect bogus news from several social media sites, the results were erroneous because of the fraudulent source.

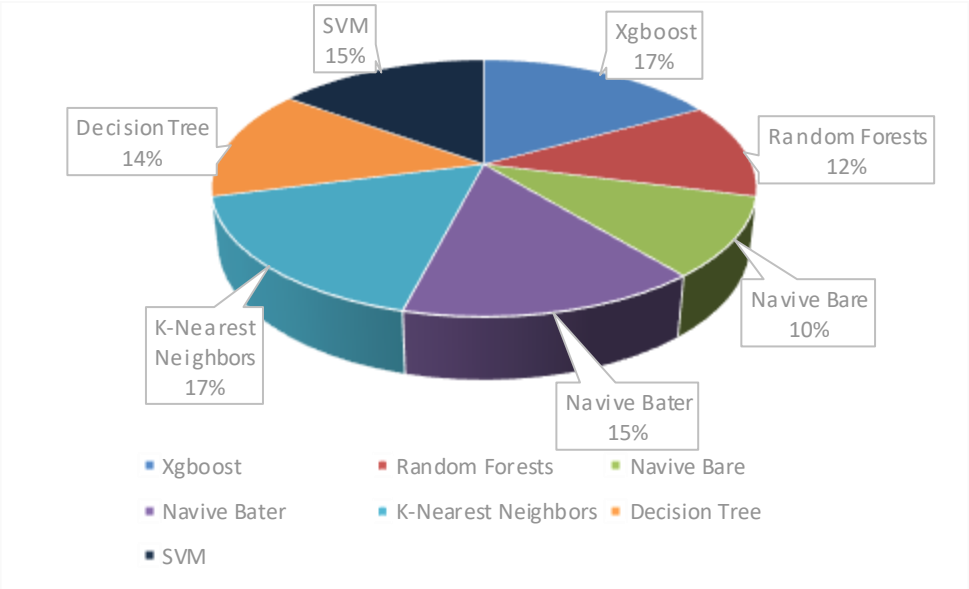

Figure 3. Results of each algorithm's accuracy

They used Kaggle to get their data, which had an accuracy rate of 74.5% on average. Twitter spammers were found using the naive bayes technique, which has an accuracy value of

70% to 71.2%. After trying with different increases, they were 76% accurate. They use three well-known methods in their research: Neural Web, Naïve Bayes, and Support Vector Machines (SVM). Naïve Bayes has a 96.08% accuracy rate in identifying fraudulent messages. The neural web and machine vector (SVM) both had 99.9% accuracy. In order to generate final results that were up to 8% more accurate than the first ones, they combined KNN and random forests with a mixed false message observation dummy. They investigated the use of eight supervised machine learning categories in the Twitter dataset while focusing on fake news related to the 2012 West Germanic elections. Additionally, they believe that the information set that has an 88% F score is the best fit for the decision tree method. When holding a straight SVM workbook and a unigram, which strategy generated the most accuracy? Use an N-gram assessment and a fake observation dummy. The tall accuracy is 92%. Furthermore, they think that the data set that has an 88% F score is the best fit for the decision tree method. When holding a straight SVM workbook and a unigram, which strategy generated the most accuracy? Use an N-gram assessment and a fake observation dummy. The tall accuracy is 92%.When holding a straight SVM workbook and a unigram, which strategy generated the most accuracy? Use an N-gram assessment and a fake observation dummy. A 92% The forecast clarity was between 70 and 76%, and the Naïve Bays algorithm was employed in the majority of research articles. They mostly used a qualitative evaluation approach, focusing on attitude surveys, word recurrence repetition, and titles. This was discovered with the use of our previous study abstract and system analysis. As a quantitative method based on the incorporation of numerical statistical data as quality, we also suggest including POS textual examination into our approach. We believe that the procession outcomes can be greatly improved by using random forests and improving their quality. The characteristics that we recommend including in our data gathering are the following metrics: Type/token ratio, total words (tokens), total unique words (types), adjectives, nouns, prepositions, sentences, characters, average word length, and so forth.

References

Ahmed A A A, Aljabouh A, Donepudi P K and Choi M S (2021). Detecting Fake News Using Machine Learning: A Systematic Literature Review. *Journal of Media Research*, 24(2), 112-128.

Ahmed H, Traore I. and Saad S (2017). Detecting Opinion Spams and Fake News Using Text Classification. *Security and Privacy*, 1(1), e9.

Collins M et al. (2021). Emerging Strategies in Fake News Detection: Surveys and Challenges. *Journal of Media Research*, 23(1), 45-67.

Goyat R, Goyat S and Sharma D (2024). Fake News Identification Using Machine Learning. *Journal of Media Research*, 25(1), 90-105.

Kaliyar R K, Goswami A, Narang P and Sinha S (2020). FNDNet—A Deep Convolutional Neural Network for Fake News Detection. *Cognitive Systems Research*, 61, 32–44.

Kansal A (2021). Fake News Detection Using POS Tagging and Machine Learning. *Journal of Applied Security Research*, 16(4), 153-171.

Kaur N et al. (2020). Multi-level Voting Model for Fake News Detection Integrating Multiple Classifiers. *Journal of Media Research*, 22(3), 215–230.

Khan,J Y, Khondaker M T I. Afroz S, Uddin G and Iqbal A (2019). A Benchmark Study of Machine Learning Models for Online Fake News Detection. *Journal of Media Research*, 21(2), 112-127.

Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., and Akbar, M. (2020). Fake News

Detection Using Deep Learning Models: A Novel Approach. *Transactions on Emerging Telecommunications Technologies*, 31(2), e3767.

Pathan H B, Bhavsingh M, Alipanahi B and Hanke A (2023). A Machine Learning-Based Approach for Detecting Fake News in Online Media. *International Journal of Computer Engineering in Research Trends*, 10(1), 22-35.

Sahoo S et al. (2021). Deep Learning-Based Fake News Detection on Facebook Using LSTM. *Journal of Media Research*, 23(4), 300-315.

Singhal A. et al. (2020). Spot Fake+: Multimodal Transfer Learning for Fake News Detection. *Journal of Media Research*, 22(4), 190-207.

Wang W Y et al. (2018). Event Adversarial Neural Networks for Multi-modal Fake News Detection. *Journal of Media Research*, 19(3), 145-157.

Zhou X et al. (2020). SAFE: Similarity-Aware Multi-Modal Fake News Detection. *Journal of Media Research*, 22(1), 50-65.

Ozbay F A and Alatas B (2020). Fake News Detection Within Online Social Media Using Supervised Artificial Intelligence Algorithms. *A: Statistical Mechanics and Its Applications*, 540, 123174.

Ozbay F A and Alatas B (2021). Adaptive Salp Swarm Optimization Algorithms for Novel Fake News Detection Model. *Multimedia Tools and Applications*, 80(26), 34333–34357.