

A Comparative Analysis of Machine Learning Classifiers for Fake News Detection using NLP

Arpita Priya, School of Engineering and Technology Sharda University, Greater Noida, UP, India
2023242913.arpita@ug.sharda.ac.in

Aanchal Vij, School of Engineering and Technology Sharda University, Greater Noida, UP, India aanchal.vij@sharda.ac.in

Abstract—The swift propagation of fake news through social media and other digital outlets represents a significant problem for information integrity and public trust in the media. We present a comparative study of machine learning models for detecting fake news in an automated fashion with a focus on a Logistic Regression, Decision Tree, Gradient Boosting and Random Forest classifiers. The paper details a straightforward methodology employing Natural Language Processing (NLP) approaches to preprocess and transform textual data: cleaning the text, removing stop words, and stemming before applying the Term Frequency - Inverse Document Frequency (TF-IDF) method for vectorization in machine learning models. Once trained on a balanced dataset of real and fake news articles, we then report comparative performance of these classifiers using key metrics including accuracy, precision, and recall, and show the results of distinguishing real from fake news. This paper details applied, interpretable, and scalable work to combat mis- and disinformation and fake news; and offer a foundation for future work or papers employing higher order techniques and datasets.

Index Terms—Fake News Detection, Machine Learning, Logistic Regression, Natural Language Processing, TF-IDF, Misinformation, Text Classification

1. Introduction

The unprecedented amount of information shared via social networking platforms has made it increasingly complicated for people and organizations to tell the difference between factual and fabricated information. This has societal implications for political conversation, public health, and social stability [4], [12], [16]. This has brought an urgency to researching the development of automated systems for fake news detection [13], [21].

Manual fact checking is invariably slow and cannot match the volume of new information that is generated daily. In response, the scientific community has begun assessing computational approaches to detection, which typically employ Artificial Intelligence (AI) and Machine Learning (ML) in order to classify and analyze news articles automatically [4], [5], [15], [34]. This work will assume a supervised machine learning approach, comparing four distinct models: Logistic Regression, Decision Trees, Gradient Boosting, and Random Forest [13]. While other machine learning methods, such as deep learning architectures, can yield excellent results [19], [20], [23], we felt that starting the development process with a baseline comparison of these proven methods was important. Logistic Regression is well known for its balanced trade-off of speed and interpretability, while tree-based ensemble methods like Random Forest and Gradient Boosting are often recognized for their high accuracy. This paper will research and accomplish the following:

- Outline the process for developing a complete end-to-end text-based fake news detection system.
- Construct and train four machine learning classifiers founded on widely used NLP techniques.
- Assess the comparative accuracy of the model using common metrics.
- Design a scalable, feasible solution to mitigate the issues caused by misinformation.

The outcomes of this research are to be offered as useful guides to practitioners, academics and organizations trying to make efficacious, transparent tools to impede the spread of fake news.

2. Literature Review

The phenomenon of fake news detection has emerged as a popular and developing area of inquiry, utilizing techniques from computer science, linguistics, and the social sciences. Significant findings have emerged around the use of different machine learning algorithms to identify fake news, studies demonstrating the successful use of popular, well-understood algorithms as well as newer deep learning models [4], [6].

In the vantage of previous research have examined various angles to the problem of detecting fake news, primarily grouping these cases by the type of features used. There are two types of features typically used in research on fake news detection : 1) content features, which typically describe characteristics of story content as it relates to topics such as linguistic style, syntax and semantics and 2) social context features, which typically describe characteristics based on a user's profile or the way in which a piece of information propagates across social

networks. [8], [24], [32]. Our approach utilizes both content and context features, emphasizes content features, primarily the text of news articles themselves since it is the more common and effective approach for fake news detection. Several systematic reviews have explored the efficacy of multiple classifiers. Alnabhan and Branco reviewed classifiers and concluded that some deep learning models, including convolutional neural networks and recurrent neural networks, often produce the best performance, achieving accuracies as high as 99.9% on certain datasets. However, these models are also often more computationally demanding, and have less interpretability than traditional machine learning models (e.g., logistic regression, naive bayes, support vector machines) that are still widely considered due to their ease of use, speed of implementation, and past success in learning tasks [3], [9], [29]. Due to its balance of performance and interpretability, we have opted for logistic regression as a starting model for our task. Foundational work in information retrieval, such as Baeza-Yates and Ramos, has developed a body of research from which we can leverage text-based feature extraction methods, such as TF-IDF, that support our work. [14], [25].

3. Available Datasets

The greatest hurdle is to establish a dataset that can assess the various techniques of fake news identification. The fake news detection dataset should produce truth factor. Various datasets collected from social media domains can assess the three main types of techniques and algorithms of fake news identification.

- 1) **FakeNewsNet Corpus [30]:** The dataset is composed of news samples collected from fact checking domains sites such as PolitiFact and GossipCop. Fake News Net has multiple dimensions of data that could be new articles, tweets, retweets, user timeline tweets, and user profiles.
- 2) **LIAR [33]:** Although the dataset is publicly available, it contains human labeled short statements extracted from POLITIFACT.COM. The short statements were collected from Facebook, Twitter, radio and television interviews and at political speeches.
- 3) **BuzzFeed-Webis Corpus:** The corpus contains 1627 articles including left wing, right wing and mainstream articles, and all publishers received Facebooks blue check mark indicating authenticity within its networks.
- 4) **BS Detector Corpus or Kaggle Fake News Dataset.:** The datasets extension identifies if there are links on a given webpage or not for use in identifying unreliable sources of information. The corpus of metadata consists of various online sites. The posts or links have been collected through the “BS detector” feature.
- 5) **FNC1 dataset [27]:** It is a dataset aimed at stance detection that contains article pairs and 75,385 labeled headlines; in each labelled headline in the dataset, a statement is paraphrased. The article pairs are labelled as agree, disagree, unrelated, and discussed.
- 6) **CredBank dataset [24]:** It contains millions of tweets that are divided into 1049 real-world situations.
- 7) **Snopes Corpus:** It has used the Snopes corpus, which has content sampled from the Snopes website, and has true claims and false claims as well.

4. Methodology

Our fake news detection system was built on a standard supervised machine learning pipeline, which involves the following key steps: (1) data collection, (2) preprocessing, (3) feature extraction, (4) model training, and (5) evaluation. This section describes the architecture and steps used to deploy our model.

A. System Architecture

The system architecture for our model to detect fake news follows a traditional machine learning pipeline for the task of text classification.

As shown in the diagram, the process begins with raw data, and then the data goes through a series of preprocessing steps. The cleaned up data is then vectorized, which is a fundamental step to convert text data to a numerical format that can be read by a machine learning model. The vectorized data is then input into the Logistic Regression model to begin training the model to classify the articles as real or fake.

B. Data Collection and Preprocessing

The model was trained using a public dataset that included a sample of true and fake news articles. The dataset was preprocessed for consistency and model performance, which included:

- **Text Cleaning:** Removing special characters, numbers, and punctuation.

- **Lowercasing:** Lowercasing all text content for data consistency.
- **Tokenization and Stop-word Removal:** Splitting text into individual words (tokens) and removing common words (e.g., the, is, a) that do not carry significant value.
- **Stemming:** Lemmatizing words (e.g., running becomes run) to both further decrease vocabulary size and treat similar words equally.

C. *Feature Extraction*

Once the text was preprocessed, the data was converted to numerical vectors via the Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer. TF-IDF is a statistical measure that determines how important a word is to a document in relation to a corpus. Hence, it assigns larger weights to words that are common in a specific document while being rare in the entire dataset. It filters and places importance on keywords.

D. *Model Training and Evaluation*

Then we split the dataset into training and testing sets to train and evaluate the model. A Logistic Regression model was trained using the TF-IDF vectorized training data: Logistic Regression, Decision Tree Classifier, Gradient Boosting Classifier and Random Forest Classifier. After training, each model was evaluated using the testing set and several metrics to give a complete view of its effectiveness skills [26]:

- **Accuracy:** The number of news articles assigned to the correct class (either real or fake).
- **Precision:** The number of fake news articles assigned the class of fake by the model out of all the articles assigned to the fake news class. **Recall:** The number of fake news articles that are true from the dataset out of all true fake news articles.

These metrics are particularly important when evaluating classifier performance on imbalanced datasets, where simply relying on accuracy can lead to misleading conclusions. [19].

5. *Results and Findings*

Our trained machine learning models all performed exceedingly well in classifying fake news articles, with the ensemble methods showing a slight edge in accuracy. As shown in Table II, the models attained the following accuracies: Logistic Regression (98.8%), Random Forest (98.9%), Decision Tree (99.5%), and Gradient Boosting (99.6%).

The confusion matrices and classification reports provide a more detailed look at the results. We see that all tested models (LR, GB, RF) have very high precision and recall for both classes. This means the models are not only overall accurate but are also adept at predicting fake news with relatively few false positives and false negatives.

6. *Evaluation Metrics*

Evaluation metric used to assess the performance of fake news classifiers include AUC (Area under curve), F1, Recall, and Accuracy. To calculate these metrics we need the number of true positive (trueP), true negative (trueN), False positive (falseP), and False negative(falseN).

Performance metric formulas are presented in Table 1.

TABLE 1: FORMULAE FOR PERFORMANCE METRICS

Metric	Formula
Precision	$\frac{\text{trueP}}{(\text{trueP} + \text{falseP})}$
Recall	$\frac{\text{trueP}}{(\text{trueP} + \text{falseN})}$
F1 Score	$\frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})}$
Accuracy	$\frac{(\text{trueP} + \text{trueN})}{(\text{trueP} + \text{falseP} + \text{falseN} + \text{trueN})}$

True Positive(trueP): number of fake news pieces assigned the class of fake news.

True Negative(trueN): number of true news pieces assigned the class of true news.

False Negative(falseN): number of true news pieces as- signed the class of fake news.
 False Positive(falseP): number of fake news pieces as- signed the class of true news.

TABLE 2: MODEL PERFORMANCE COMPARISON

Metric	Logistic Regression	Decision Tree	Gradient Boosting	Random Forest
Accuracy	0.988	0.995	0.996	0.989
Precision	0.990	N/A*	1.000	0.990
Recall	0.990	N/A*	1.000	0.990
F1 Score	0.990	N/A*	1.000	0.990

7. Conclusion and Future Scope

This research has demonstrated that several machine learning models, particularly Gradient Boosting and Decision Trees can be effectively employed as a solution for automatically detecting fake news. Logistic Regression and Random Forest also provide robust and reliable methods for text classification, with accuracies approaching 99%. As the fake news challenge continues to grow and change, future work may consider:

- Integration of Deep Learning: More sophisticated models, such as LSTMs or Transformers, that could potentially leverage more complex semantic and contextual features [19], [20], [22].
- Multimodal Detection: Non-textual features such as images, videos, and social network propagation to create a more elaborate fake news detection system. [14], [15].
- Real-Time Deployment: A live system that could read and classify news as it spreads through social media [34] by utilizing our foundational work, we can move forward in creating more sophisticated and better strategies for responding to the increasing amounts of misleading information in the digital age.

References

- [1] S. I. Manzoor, J. Singla, and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 230-234.
- [2] M. K. Elhadad, K. F. Li, and F. Gebali, "Fake News Detection on Social Media: A Systematic Survey," in 2019 Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), 2019, pp. 1-6.
- [3] V. Rampurkar and T. D. R., "An Approach towards Fake News Detection using Machine Learning Techniques," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 3, pp. 2868-2874, 2024.
- [4] M. Q. Alnabhan and P. Branco, "Fake News Detection Using Deep Learning: A Systematic Literature Review," IEEE Access, 2024.
- [5] S. B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," in 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 436-441.
- [6] B. M. Nema and A. Al-Sherbaz, "Deep Learning for Fake News Detection: Literature Review," Al-Mustansiriyah J. Sci., vol. 34, no. 2, pp. 70-81, 2023.
- [7] F. D. C. Medeiros and R. B. Braga, "Fake News Detection in Social Media: A Systematic Review," Int. J. Hum.-Comput. St., 2020.
- [8] S. S. Nikam and R. Dalvi, "Fake News detection Using Machine Learning," in 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), 2021, pp. 1- 5.
- [9] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," Security and Privacy, vol. 1, no. 1, p. e9, Jan. 2018.
- [10] M. Albahar, "A hybrid model for fake news detection: Leveraging news content and user comments in fake news," IET Information Security, vol. 15, no. 1, pp. 1-13, 2021.
- [11] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211-236, 2017.
- [12] N. F. Baair and A. Djeflal, "Fake News detection Using Machine Learning," in 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), 2021.
- [13] R. Baeza-Yates and B. Ribeiro-Neto, Modern information retrieval, 1st ed. New York: ACM Press, 1999.
- [14] T. Chakraborty, "Multi-modal Fake News Detection," in Data Science for Fake News, Springer, 2021, pp. 85-110.
- [15] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic Deception Detection: An Introduction to a New Research Area," First Monday, vol. 22, no. 11, 2017.
- [16] P. H. A. Faustini and T. F. Covoˆes, "Fake news detection in multiple platforms and languages," Expert Systems with Applications, vol. 158, 2020.
- [17] S. Gilda, "Evaluating machine learning algorithms for fake news detection," in 2017 IEEE 15th Student Conference on Research and Development (SCORED), 2017, pp. 110-115.

- [18] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," arXiv preprint arXiv:2002.01030, 2020.
- [19] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645-6649.
- [20] C. Guo, J. Cao, X. Zhang, K. Shu, and M. Yu, "Exploiting emotions for fake news detection on social media," arXiv preprint arXiv:1903.01728, 2019.
- [21] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," Trans. Emerg. Telecommun. Technol., vol. 31, no. 2, p. e3767, Feb. 2020.
- [22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.
- [23] T. Mitra and E. Gilbert, "Credbank: A large-scale social media corpus with associated credibility annotations," in Proceedings of the International AAAI Conference on Web and Social Media, 2015.
- [24] J. Ramos, "Using tf-idf to determine word relevance in document queries," in Proceedings of the First Instructional Conference on Machine Learning, 2003, pp. 133-142.
- [25] J. C. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Explainable machine learning for fake news detection," in Proceedings of the 10th ACM Conference on Web Science, 2019, pp. 17-26.
- [26] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, "A simple but tough-to-beat baseline for the Fake News Challenge stance detection task," arXiv preprint arXiv:1707.03264, 2017.
- [27] V. L. Rubin and T. Lukoianova, "Truth and deception at the rhetorical structure level," Journal of the Association for Information Science and Technology, vol. 66, no. 5, pp. 905-917, 2015.
- [28] B. Scholkopf and A. J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, 2018.
- [29] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media," arXiv preprint arXiv:1809.01286, 2018.
- [30] K. Shu, D. Mahudeswaran, S. Wang, R. Zafarani, and H. Liu, "defend: Explainable fake news detection," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, 2019, pp. 395-405.
- [31] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profile for fake news detection," arXiv preprint arXiv:1904.13355, 2019.
- [32] W. Y. Wang, "Liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint arXiv:1705.00648, 2017.
- [33] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional neural networks for fake news detection," arXiv preprint arXiv:1809.01286, 2018.
- [34] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, "Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads," in International AAAI Conference on Web and Social Media, 2015.