International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Fake News Detection Using Machine Learning And Natural Language Processing: A Review Of Text-Based Classification Pipelines

Bangar Piyush Rajesh[1]; Prof. S. K. Said[2]; Prof. M. C. Kolse[3]
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—Social media, websites and online portals have become major sources of news, but they also allow false and misleading information to spread rapidly, creating confusion and social harm. Fake news detection aims to identify such content automatically so that users and platforms can respond quickly. This paper reviews machine learning and natural language processing approaches to detecting fake news from the text of articles and headlines. It covers how fake news is defined and why it spreads, benchmark datasets such as LIAR, ISOT and FakeNewsNet, preprocessing steps including cleaning, tokenization and stop-word removal, feature extraction with bag-of-words, n-grams and TF-IDF, and classification with Logistic Regression, Naive Bayes and Support Vector Machines, as well as deep models based on BERT. A practical five-layer pipeline that can be implemented with Python libraries such as Pandas, NumPy, Scikit-learn and NLTK is proposed, together with research gaps, objectives and a comparison of algorithms. Limitations of text-only detection, dataset bias and the need for explainable, regularly updated models are discussed. The review concludes that a well-built text-classification pipeline is an effective first line of defence and a strong foundation for richer detection systems.

Index Terms—Fake news detection, machine learning, natural language processing, TF-IDF, logistic regression, naive Bayes, support vector machine, misinformation

I. INTRODUCTION

People increasingly read news through social media feeds and messaging apps rather than newspapers. Anyone can publish, and content can be shared thousands of times within hours. Vosoughi *et al.* [1] studied the spread of verified true and false stories on Twitter and found that falsehoods spread farther and faster than the truth, largely because people find novel claims more shareable.

Manual fact-checking is careful but slow and cannot keep up with this volume. Automatic detection uses machine learning to flag suspicious articles for review or warning. Shu *et al.* [2] framed fake news detection as a data mining problem involving news content, social context and the spread of information.

This paper focuses on text-based detection, the most accessible approach for a student project, and reviews datasets, features and classifiers before proposing a practical pipeline.

II. THEORETICAL BACKGROUND

2.1 Types Of False Information

Fig. 1 distinguishes satire, misleading content, fabricated content, manipulated content and false context. Detection methods differ in whether they analyse writing style, factual claims, the source or the way content spreads.



Fig. 1. Types of false information and detection perspectives.

2.2 Text Representation

Bag-of-words counts word occurrences; n-grams capture short phrases; and TF-IDF weights terms that are frequent in one document but rare overall. Transformer models such as BERT [3] produce contextual embeddings that capture meaning beyond individual words.

III. SURVEY OF EXISTING APPROACHES

Wang [4] released LIAR, a dataset of short political statements labelled by fact-checkers into six truthfulness levels, and showed that adding speaker metadata to text improved classification. Ahmed *et al.* [5] compiled the ISOT dataset of real and fake articles and found that n-gram TF-IDF features with linear classifiers such as linear SVM performed best among several machine learning methods.

Pérez-Rosas *et al.* [6] built datasets of fake and legitimate news across domains and showed that linguistic features, including readability, punctuation and psycholinguistic categories, distinguish fake news, but that models degrade

© 2026 IJSRET

when applied to new domains. Potthast *et al.* [7] analysed writing style and found that hyperpartisan news is easier to identify by style than fake news itself, a warning that style alone is not enough.

Shu *et al.* [8] released FakeNewsNet, which pairs article content with social context such as user engagement. Kaliyar *et al.* [9] proposed FakeBERT, combining BERT with convolutional layers, and reported strong accuracy on a real-world dataset. Zhou and Zafarani [10] surveyed the field and grouped methods into knowledge-based, style-based, propagation-based and source-based detection. Table I summarizes these works.

TABLE I Summary of Literature Survey

Sr. No.	Ref.	Year	Methodology	Findings	Research Gap
1	[4]	2017	LIAR dataset; hybrid CNN	Metadata improves accuracy	Short statements only
2	[5]	2017	N-gram TF-IDF with ML classifiers	Linear SVM performed best	Single dataset
3	[1]	2018	Large-scale diffusion analysis	False news spreads faster than true	Twitter only
4	[6]	2018	Linguistic feature analysis	Style features identify fake news	Poor cross-domain transfer
5	[7]	2018	Stylometric analysis	Hyperpartisan style detectable	Style alone insufficient
6	[8]	2020	FakeNewsNet repository	Content plus social context	Mostly US politics
7	[10]	2020	Survey of fake news detection	Four-perspective taxonomy	Survey only
8	[9]	2021	FakeBERT (BERT + CNN)	High accuracy on real data	Compute-intensive

IV. COMPARATIVE STUDY

Table II compares the classifiers. Naive Bayes is fastest, Logistic Regression and linear SVM are strong on TF-IDF features, and BERT-based models achieve the best understanding of context at higher computational cost.

Classifier	Features	Speed	Accuracy
Naive Bayes	TF-IDF / counts	Very fast	Good
Logistic Regression	TF-IDF	Fast	High
Linear SVM	TF-IDF	Fast	High
Random Forest	TF-IDF + style	Medium	Medium–High
BERT-based	Contextual embeddings	Slow	Very high

TABLE II Comparison of Classifiers

V. RESEARCH GAPS

- Models trained on one news domain or period often fail on new topics and events.
- Most datasets are in English; Indian languages and code-mixed text are under-represented.
- Text-only models ignore images, sources and sharing patterns that carry useful signals.
- Many systems output a label without explaining which words or claims caused it.
- Fake news evolves quickly, but models are rarely updated.

VI. AIMS AND OBJECTIVES

- To review the nature of fake news and available datasets.
- To study preprocessing and feature extraction methods for news text.
- To compare Logistic Regression, Naive Bayes, SVM and BERT-based classifiers.
- To propose a practical, Python-based fake news detection pipeline.
- To identify limitations and future improvements.

VII. PROPOSED ARCHITECTURE

The proposed pipeline is shown in Fig. 2.

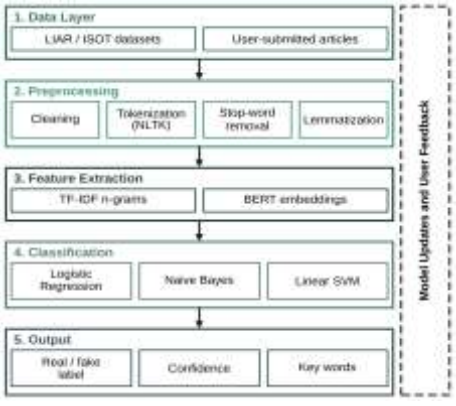


Fig. 2. Proposed text-based fake news detection pipeline.

© 2026 IJSRET

7.1 Data Layer

Labelled articles are loaded from public datasets with Pandas; new articles are submitted by users through a web form.

7.2 Preprocessing Layer

Text is lower-cased, URLs and punctuation are removed, words are tokenized with NLTK, stop words are removed and words are stemmed or lemmatized.

C. Feature Extraction Layer

TF-IDF vectors with unigrams and bigrams represent each article; optionally, BERT embeddings provide contextual features.

7.3 Classification Layer

Logistic Regression, Multinomial Naive Bayes and linear SVM are trained with Scikit-learn and compared; the best model is saved.

E. Output Layer

The user sees a real or fake prediction with a confidence score and the most influential words, plus a reminder to verify with trusted sources.

VIII. PROCESS FLOW AND EVALUATION

Fig. 3 shows the workflow. TF-IDF for term t in document d is

$$tfidf(t,d) = tf(t,d) \cdot \log(N / (1 + df(t))) \quad (1)$$

and Logistic Regression estimates the probability of fake news as

$$P(fake \mid x) = 1 / (1 + e^{-(w^T x + b)}) \quad (2)$$

Models are compared with accuracy, precision, recall and F1-score using cross-validation.

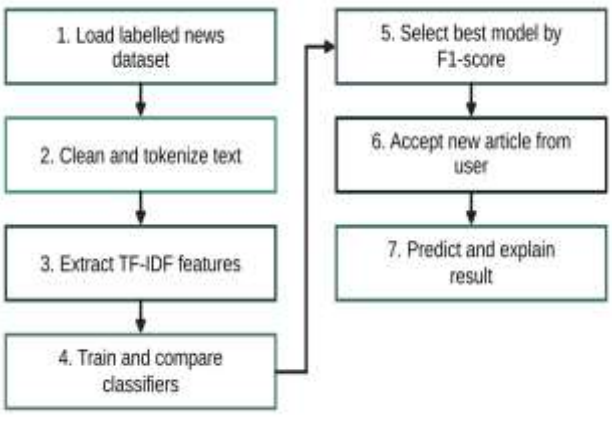


Fig. 3. Workflow of the fake news detection system.

IX. LIMITATIONS AND FUTURE SCOPE

Table III lists challenges. Future extensions include multilingual models for Indian languages, combining text with source credibility and propagation features, linking to fact-checking databases, and browser extensions that warn users before they share.

TABLE III Challenges and Mitigation

Challenge	Mitigation
New topics and events	Regular retraining; fact-check links
Indian languages	Multilingual models (e.g., mBERT)
Unexplained labels	Show influential words
Dataset bias	Balanced, diverse datasets
Over-blocking	Warnings instead of removal

X. CONCLUSION

Automatic fake news detection helps users and platforms respond to misinformation at scale. The reviewed literature shows that TF-IDF features with linear classifiers form a strong, efficient baseline, while transformer models capture deeper context. A simple Python pipeline can classify news articles quickly, but it should be treated as decision support, explained to users and regularly updated as misinformation tactics change.

10.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

REFERENCES

1. S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
2. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017.
3. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
4. W. Y. Wang, ““Liar, liar pants on fire”: A new benchmark dataset for fake news detection,” in Proc. 55th Annu. Meeting ACL, 2017, pp. 422–426.

© 2026 IJSRET

5. H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using n-gram analysis and machine learning techniques,” in Proc. Int. Conf. Intell., Secure, Dependable Syst. Distrib. Cloud Environ. (ISDDC), 2017, pp. 127–138.
6. V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” in Proc. 27th Int. Conf. Comput. Linguist. (COLING), 2018, pp. 3391–3401.
7. M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, “A stylometric inquiry into hyperpartisan and fake news,” in Proc. 56th Annu. Meeting ACL, 2018, pp. 231–240.
8. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media,” Big Data, vol. 8, no. 3, pp. 171–188, 2020.
9. R. K. Kaliyar, A. Goswami, and P. Narang, “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach,” Multimed. Tools Appl., vol. 80, pp. 11765–11788, 2021.
10. X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories, detection methods, and opportunities,” ACM Comput. Surv., vol. 53, no. 5, Art. no. 109, 2020.

© 2026 IJSRET