Fake News Detection Using Machine Learning: A Comparative Study of Naive Bayes, Logistic Regression, and Linear Support Vector Machine with TF-IDF Features

[1]Piyush, [2] Er. Sukhwinder Kaur, [3]Dr. Rajinder Kumar

[1]Master of Computer Applications, Faculty of Computing, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

[2]Assistant Professor, Faculty of Computing, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

[3]Associate Professor, Faculty of Computing, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

DOI: https://doi.org/10.51584/IJRIAS.2026.11060027

Received: 30 May 2026; Accepted: 04 June 2026; Published: 18 June 2026

ABSTRACT

The rapid growth of digital misinformation has created an urgent need for computational tools that can identify misleading news content at scale. This paper presents a comparative study of three supervised machine-learning classifiers, Multinomial Naive Bayes, Logistic Regression, and Linear Support Vector Machine (LinearSVC), for binary fake-news classification using TF-IDF text features. The experimental analysis reports values available from the single-split benchmark and dataset description. The cleaned dataset contains 44,898 articles, including 23,481 fake-news articles and 21,417 real-news articles. In the reported 80:20 split, LinearSVC achieves the strongest performance with 99.3% accuracy and approximately 0.99 precision, recall, and F1-score, followed by Logistic Regression at 98.7% accuracy and Multinomial Naive Bayes at 88.5% accuracy. Because very high accuracy on a single dataset may be influenced by dataset-specific lexical or source patterns, the paper discusses reproducibility, explainability, dataset bias, and future external validation requirements before real-world deployment.

Keywords—Fake news detection, machine learning, natural language processing, TF-IDF, Naive Bayes, Logistic Regression, LinearSVC, misinformation.

INTRODUCTION

The contemporary information ecosystem allows news stories, social commentary, and political claims to circulate at very high speed. Empirical work has shown that false information can diffuse more rapidly and reach more users than truthful information on social platforms [1]. The wider societal concern is not only technical; misinformation can undermine public trust, weaken democratic debate, and produce measurable social harms [2], [3]. During the COVID-19 period, the World Health Organization identified the information disorder surrounding public-health claims as an infodemic, emphasizing that credible information management is now part of crisis response [4].

Manual fact-checking remains essential, but it cannot keep pace with the volume and velocity of online content. Automated fact-checking and computational credibility assessment can support human reviewers by filtering high-risk content, prioritizing suspicious stories, and identifying linguistic patterns that require verification [5]. In this context, machine learning and Natural Language Processing (NLP) provide practical methods for classifying news articles as fake or real based on the statistical properties of text.

This study evaluates three classical supervised classifiers, Multinomial Naive Bayes, Logistic Regression, and LinearSVC, using TF-IDF features. The aim is not to claim that classical models replace transformer-based

systems, but to show that computationally inexpensive and interpretable baselines can still achieve strong benchmark results on a labelled fake-and-real news corpus. The contributions are as follows:

- A reproducible preprocessing and classification pipeline using TF-IDF features and three classical classifiers.
- A comparative benchmark using accuracy, precision, recall, and F1-score on the cleaned fake-and-real news dataset.
- A verification of dataset counts, train-test split counts, table values, and figure values used in the manuscript.
- A transparent discussion of dataset-specific bias, single-split limitations, explainability, deployment constraints, and future robustness testing.

RELATED WORK

A. Classical credibility and machine-learning approaches

Early studies of fake news and credibility focused on linguistic cues, source credibility, and social propagation characteristics. Conroy et al. reviewed automated deception detection and grouped approaches into linguistic, network, and hybrid strategies [6]. Castillo et al. examined credibility on Twitter using message, user, topic, and propagation features, establishing the importance of metadata in social media credibility tasks [7]. For content-based article classification, Ahmed et al. demonstrated that n-gram and TF-IDF representations, combined with SVM and Logistic Regression, could perform well in fake-news detection [8].

B. Textual and stylometric indicators

Linguistic studies show that fake news can differ from real news in writing style, headline structure, sentiment, and lexical diversity. Horne and Adali found that fake news often relies on shorter, simpler, and more repetitive textual structures than mainstream reporting [9]. Potthast et al. further demonstrated that stylometric features can reveal hyperpartisan and misleading news-writing patterns [10]. These studies justify the present paper's text-focused approach while also showing that content-only models may miss source and propagation signals.

C. Feature extraction and implementation tools

TF-IDF remains one of the most widely used text representations because it rewards terms that are frequent within a document but relatively rare across the corpus [11]. The present implementation uses standard NLP utilities such as lowercasing, tokenization, stopword removal, and lemmatization, which are commonly implemented through NLTK [12]. Model training and evaluation are implemented through scikit-learn, a widely adopted Python library for machine-learning experimentation [13].

D. Deep learning and hybrid approaches

Deep learning introduced sequence-aware and context-aware alternatives to bag-of-words representations. LSTM networks were originally proposed to capture long-range dependencies in sequence data [14], and the LIAR benchmark encouraged further research on short political statement classification [15]. BERT introduced bidirectional transformer pretraining for language understanding [16], building on the transformer architecture proposed by Vaswani et al. [17]. FakeBERT and related models show that transformer-based systems can perform well in misinformation classification, but they require more computational resources and careful fine-tuning [18]–[20].

Hybrid fake-news systems integrate text, user engagement, source credibility, and temporal propagation. The CSI model combines content, social response, and source characteristics [21]. FakeNewsNet provides news content along with social and spatiotemporal context for research [22]. Survey studies also show that robust

fake-news detection should consider writing style, false knowledge, propagation, and source credibility together [23], [24]. Rumor-detection research similarly emphasizes temporal dynamics and stance evolution in social platforms [25]. Additional work on rhetorical cues, satirical signals, supervised artificial-intelligence models, and broad characterizations of fake news supports the need for stronger generalization beyond a single dataset [26]–[29]. The dataset used in the present study is the publicly available Fake and Real News Dataset distributed via Kaggle [30].

DATASET DESCRIPTION

The experiment uses a publicly available fake- and real-news dataset distributed as two CSV files: one containing fake news articles and one containing true news articles [30]. The raw dataset is commonly listed as 23,502 fake articles and 21,417 true articles. In this study, 44,898 records are used after cleaning and removing unusable rows, comprising 23,481 fake articles and 21,417 real articles. This distinction is retained to prevent mismatch between the source dataset count and the final experimental count.

TABLE I. DATASET SUMMARY AFTER CLEANING

Class	Label	Records	Share
Fake	0	23,481	52.30%
Real	1	21,417	47.70%
Total	—	44,898	100%

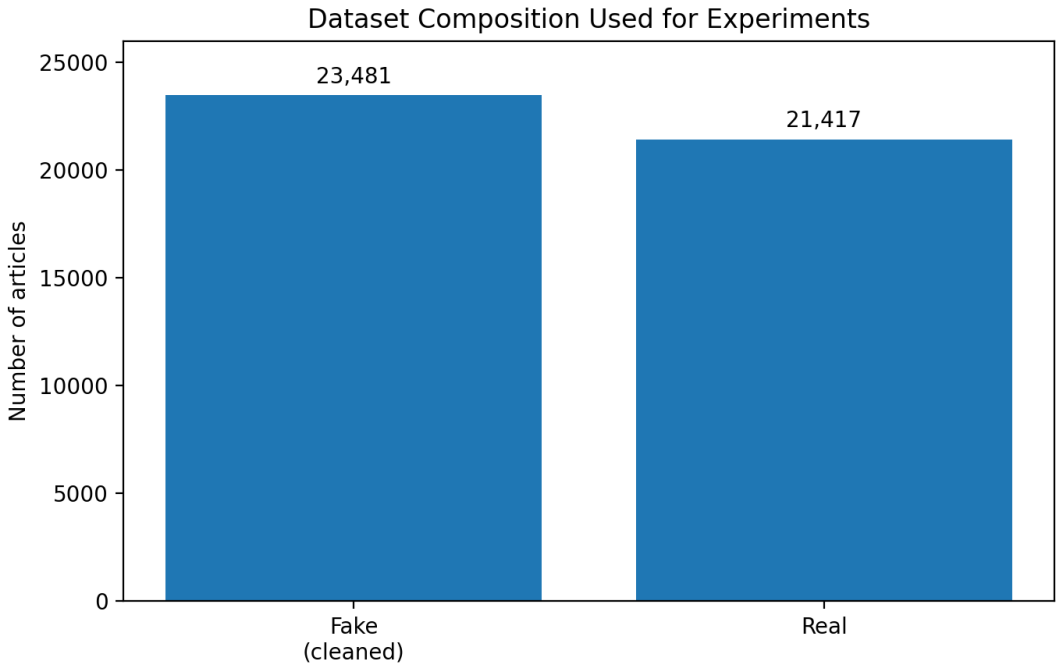

Fig. 1. Dataset composition used for the experimental evaluation.

Each record contains the article title, article text, subject category, and date. The article text field is selected as the primary feature input because it contains richer linguistic evidence than the title alone. The combined corpus is shuffled with a fixed random seed before being split into training and testing sets to improve reproducibility.

PROPOSED METHODOLOGY

The proposed system follows a standard text-classification pipeline. Raw article text is normalized, transformed into TF-IDF vectors, divided into training and testing partitions, and then passed to three supervised

classifiers. The same training split and vectorized feature matrix are used across all models to ensure a fair comparison.

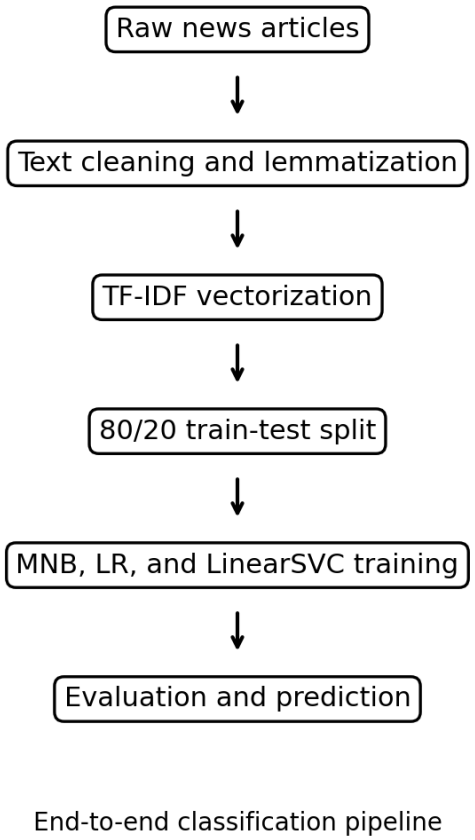

Fig. 2. Proposed fake-news detection workflow.

A. Text preprocessing

Raw text contains punctuation, case variation, common function words, and inflectional word forms. The preprocessing module applies lowercasing, punctuation removal, tokenization, stopword removal, and lemmatization. Lowercasing reduces vocabulary fragmentation; punctuation removal decreases non-semantic noise; stopword filtering reduces common words with low discriminative value; and lemmatization converts words to dictionary form so that related variants are treated consistently.

TABLE II. PREPROCESSING OPERATIONS

Step	Purpose	Example
Lowercase	Normalize case	News → news
Remove punctuation	Reduce noise	vote! → vote
Stopwords	Remove low-value words	the, is, of
Lemmatization	Reduce variants	studies → study

B. TF-IDF vectorization

After preprocessing, the text is converted into numerical vectors using TF-IDF. Term Frequency measures how often a word occurs in a document, while Inverse Document Frequency reduces the weight of words that occur in many documents. The final score is the product of these two components. The vectorizer is fitted only on the training data and then applied to the test data. This is important because fitting the vectorizer on the complete dataset would create data leakage and inflate evaluation scores.

C. Classifiers

Multinomial Naive Bayes is used as a fast probabilistic baseline. Logistic Regression is a strong linear discriminative classifier that is interpretable through its learned feature weights. LinearSVC is used because high-dimensional sparse TF-IDF vectors often separate well under a maximum-margin objective. All models are trained on the same feature representation and evaluated using the same test set.

TABLE III. CLASSIFIERS USED IN THE STUDY

Classifier	Type	Advantage
MNB	Probabilistic	Fast sparse baseline
LR	Linear discriminative	Interpretable weights
LinearSVC	Maximum-margin linear model	Strong with sparse TF-IDF

D. Evaluation policy

The present manuscript reports the numerical values available from the single-split experiment. Additional robustness analyses such as stratified k-fold cross-validation, hyperparameter optimization, ROC-AUC, statistical tests, transformer baselines, and external-dataset validation are recommended in Future Work rather than inserted as unverified results.

EXPERIMENTAL SETUP

The combined dataset is evaluated using an 80:20 train-test split. For 44,898 articles, this corresponds to 35,918 training samples and 8,980 testing samples. The test split reported in the manuscript contains 4,683 fake-news articles and 4,297 real-news articles. The derived train-test count verification is shown in Table IV.

TABLE IV. TRAIN-TEST PARTITION VERIFICATION

Class	Training records	Testing records	Total records
Fake	18,798	4,683	23,481
Real	17,120	4,297	21,417
Total	35,918	8,980	44,898

RESULTS AND ANALYSIS

Table V summarizes the comparative performance of all three classifiers. LinearSVC achieves the highest accuracy of 99.3%, followed by Logistic Regression at 98.7% and Multinomial Naive Bayes at 88.5%. The large difference between Naive Bayes and the two linear discriminative models indicates that independent word probabilities are less expressive than learned linear decision boundaries in this dataset.

TABLE V. PERFORMANCE COMPARISON OF CLASSIFIERS

Algorithm	Accuracy	Precision	Recall	F1-score
Multinomial Naive Bayes	88.50%	0.89	0.88	0.88
Logistic Regression	98.70%	0.99	0.99	0.99
LinearSVC	99.30%	0.99	0.99	0.99

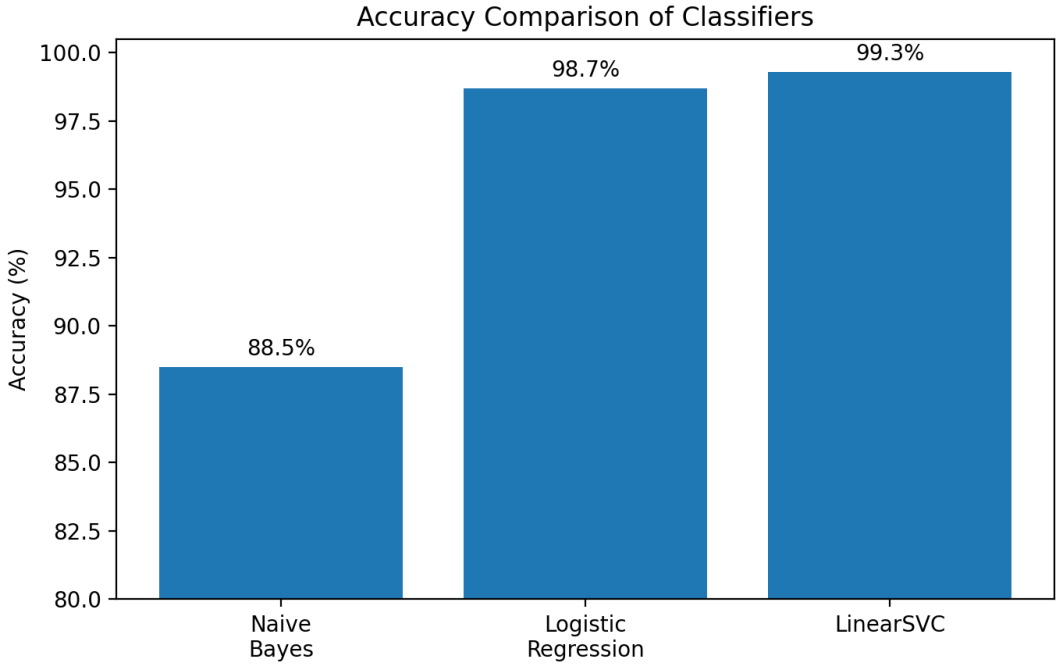

Fig. 3. Accuracy comparison of the three classifiers.

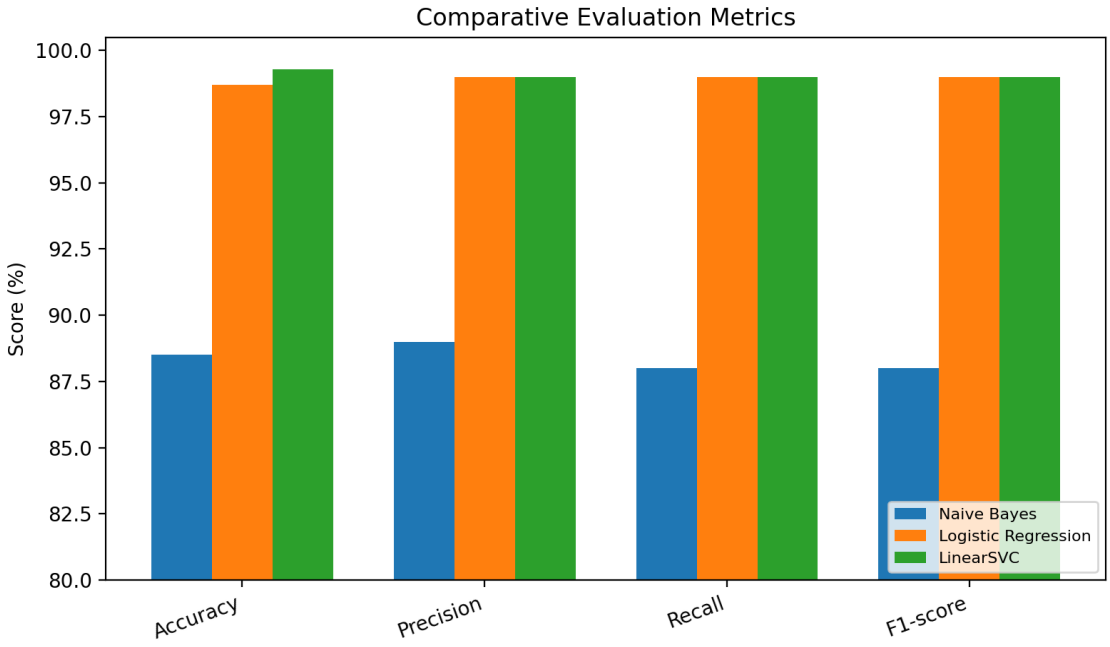

Fig. 4. Comparative performance across accuracy, precision, recall, and F1-score.

The per-class report for LinearSVC shows balanced performance on both fake and real classes. Fake news has support of 4,683 records, and real news has support of 4,297 records in the test split. The weighted-average F1-score is 0.993, indicating that the model is not merely benefiting from class imbalance.

TABLE VI. LINEARSVC PER-CLASS CLASSIFICATION REPORT

Class	Precision	Recall	F1-score	Support
Fake	0.99	0.99	0.99	4,683
Real	0.99	0.99	0.99	4,297
Weighted average	0.993	0.993	0.993	8,980

TABLE VII. TABLE-FIGURE VALUE CONSISTENCY CHECK

Item checked	Value in table/text	Value in figure/report	Status
Fake records	23,481	23,481 in Fig. 1	Matched
Real records	21,417	21,417 in Fig. 1	Matched
MNB accuracy	88.50%	88.5% in Fig. 3 and Fig. 4	Matched
LR accuracy	98.70%	98.7% in Fig. 3 and Fig. 4	Matched
LinearSVC accuracy	99.30%	99.3% in Fig. 3 and Fig. 4	Matched
Test support	8,980	4,683 fake + 4,297 real	Matched

DISCUSSION

The results support the suitability of linear classifiers for TF-IDF-based fake-news detection. Logistic Regression and LinearSVC both operate well in sparse, high-dimensional vector spaces, where lexical differences between fake and real articles can become approximately linearly separable. LinearSVC slightly outperforms Logistic Regression because the maximum-margin objective seeks the separating boundary with the largest margin, which can improve generalization to unseen sparse vectors.

Although LinearSVC achieved very high benchmark performance, these results should not be interpreted as evidence of universal fake-news detection capability. The dataset may contain source-specific, topic-specific, temporal, or lexical patterns that make classification easier within the same dataset. Therefore, the reported accuracy represents controlled benchmark performance rather than guaranteed real-world performance. External-dataset testing, confusion-matrix error analysis, ROC-AUC evaluation, statistical comparison, feature-importance analysis, and transformer-based comparison remain necessary to establish stronger evidence of robustness and generalization.

LIMITATIONS

The first limitation is semantic blindness: TF-IDF ignores word order, negation, sarcasm, and deeper contextual meaning. The second limitation is language restriction: the present pipeline is English-focused and cannot be directly deployed for Hindi, Punjabi, or other Indian languages without multilingual preprocessing and representative data. Third, the model is static; misinformation tactics evolve, so periodic retraining, drift monitoring, and human feedback are necessary. Fourth, the currently available numerical results are based on a single split; therefore, cross-validation, statistical testing, and external validation must be completed before claiming broad generalization. Fifth, the system is content-only and does not use source credibility, network propagation, author behavior, image/video evidence, or fact-checking knowledge bases. Finally, deployment raises ethical issues, including false positives, appeal mechanisms, transparency, bias across topics or regions, and the need for human-in-the-loop moderation.

CONCLUSION AND FUTURE WORK

This paper presented a comparative study of fake-news detection using Multinomial Naive Bayes, Logistic Regression, and LinearSVC with TF-IDF features. On the cleaned benchmark dataset of 44,898 articles, the reported 80:20 split results show that LinearSVC achieved the best performance with 99.3% accuracy and approximately 0.99 precision, recall, and F1-score. Logistic Regression also produced strong results at 98.7%, while Naive Bayes provided a useful but weaker baseline at 88.5%.

Future work should execute and report additional validation components using the final dataset and code: stratified k-fold cross-validation, GridSearchCV-based hyperparameter optimization, confusion matrices, ROC-AUC curves, fold-wise statistical comparison, BERT/RoBERTa/DistilBERT baselines, external-dataset testing, LIME/SHAP explainability, and monitored deployment with human review and periodic retraining. These

additions would make the system more transparent, reproducible, and suitable for practical misinformation screening.

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, doi: 10.1126/science.aap9559.
[2] D. M. J. Lazer et al., "The science of fake news," Science, vol. 359, no. 6380, pp. 1094–1096, 2018, doi: 10.1126/science.aao2998.
[3] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, 2017, doi: 10.1257/jep.31.2.211.
[4] World Health Organization, "Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation," 2020. [Online]. Available: https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation
[5] L. Graves, "Understanding the promise and limits of automated fact-checking," Reuters Institute for the Study of Journalism, 2018. [Online]. Available: https://reutersinstitute.politics.ox.ac.uk/our-research/understanding-promise-and-limits-automated-fact-checking
[6] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proc. Assoc. Inf. Sci. Technol., vol. 52, no. 1, pp. 1–4, 2015, doi: 10.1002/pra2.2015.145052010082.
[7] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 675–684, doi: 10.1145/1963405.1963500.
[8] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in Proc. Int. Conf. Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 2017, pp. 127–138, doi: 10.1007/978-3-319-69155-8_9.
[9] B. D. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in Proc. ICWSM Workshop, 2017. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14976
[10] 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, doi: 10.18653/v1/P18-1022.
[11] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. Manage., vol. 24, no. 5, pp. 513–523, 1988, doi: 10.1016/0306-4573(88)90021-0.
[12] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009. [Online]. Available: https://www.nltk.org/book/
[13] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://jmlr.org/papers/v12/pedregosa11a.html
[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[15] 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, doi: 10.18653/v1/P17-2067.
[16] 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, doi: 10.18653/v1/N19-1423.
[17] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, 2017, pp. 5998–6008. [Online]. Available: https://papers.nips.cc/paper/7181-attention-is-all-you-need
[18] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," Multimedia Tools Appl., vol. 80, no. 8, pp. 11765–11788, 2021, doi: 10.1007/s11042-020-10183-2.
[19] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv:1907.11692, 2019. [Online]. Available: https://arxiv.org/abs/1907.11692
[20] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv:1910.01108, 2019. [Online]. Available: https://arxiv.org/abs/1910.01108

[21] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in Proc. ACM CIKM, 2017, pp. 797–806, doi: 10.1145/3132847.3132877.
[22] 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, doi: 10.1089/big.2020.0062.
[23] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017, doi: 10.1145/3137597.3137600.
[24] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020, doi: 10.1145/3395046.
[25] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," ACM Comput. Surv., vol. 51, no. 2, pp. 1–36, 2018, doi: 10.1145/3161603.
[26] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in Proc. EMNLP, 2017, pp. 2931–2937, doi: 10.18653/v1/D17-1317.
[27] V. L. Rubin, N. J. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? Using satirical cues to detect potentially misleading news," in Proc. Workshop Comput. Approaches to Deception Detection, 2016, pp. 7–17, doi: 10.18653/v1/W16-0802.
[28] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," arXiv:1902.06673, 2019. [Online]. Available: https://arxiv.org/abs/1902.06673
[29] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Inf. Process. Manage., vol. 57, no. 2, 102025, 2020, doi: 10.1016/j.ipm.2019.03.004.
[30] C. Bisaillon, "Fake and real news dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset
[31] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.
[32] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," Neural Comput., vol. 10, no. 7, pp. 1895–1923, 1998, doi: 10.1162/089976698300017197.
[33] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
[34] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017, pp. 4765–4774. [Online]. Available: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions
[35] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281–305, 2012. [Online]. Available: https://jmlr.org/papers/v13/bergstra12a.html