A Machine Learning Approach for Fake News Detection Using NLP-Based Feature Extraction

Pratiksha Venkatesh Mane[1], Tanuja Jitendra Walunj[2], Sanskriti Ramdas Dhobale[3], Ankita Ashok Bhole[4], Ashwini Sanjay Gagare[5]

[1,2,3,4,5]*Department of Science & Computer Science MAEER's MIT Art's, Commerce and Science College, Alandi (D), Pune*

doi.org/10.64643/ijirt.208742-459

***Abstract*—The rapid growth of digital media and social networking platforms has made it easier for information to be created and distributed at a large scale. Although online platforms provide fast access to news, they also enable the rapid spread of fake news, misleading information, and fabricated content. Fake news can influence public opinion, create social unrest, damage reputations, and affect political, economic, and public-health decisions. Therefore, automated fake-news detection has become an important research area in Natural Language Processing (NLP) and machine learning. This paper presents a machine-learning approach using text preprocessing, TF-IDF representation, n-gram analysis, linguistic features, and supervised classification. Naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest can be compared using accuracy, precision, recall, F1-score, and confusion matrices. The approach is efficient and comparatively interpretable, but its reliability depends on dataset quality, labeling, changing language patterns, and external verification.**

***Index Terms*—Fake News Detection; Machine Learning; Natural Language Processing; TF-IDF; Text Classification; Naive Bayes; Logistic Regression; Support Vector Machine; Responsible AI**

I. INTRODUCTION

The internet has changed the way people create, access, and share news. Online newspapers, blogs, social-media platforms, and messaging applications allow information to reach a large audience within seconds. However, this accessibility also enables the rapid distribution of false, manipulated, and misleading information. Fake news generally refers to fabricated or misleading content presented in the form of legitimate news. It may be produced for political influence, financial gain, propaganda, entertainment, or public manipulation.

Manual verification is difficult because thousands of articles and posts are published every day. A fact-checker may need to examine the source, publication history, writing style, evidence, and factual claims before making a decision. Machine-learning systems can support this process by learning patterns from previously labeled real and fake news articles and using those patterns to classify unseen content.

Natural Language Processing is important because news articles are unstructured textual data. Preprocessing and feature extraction convert text into numerical representations that can be understood by a machine-learning algorithm. The objectives of this study are to preprocess news articles, extract meaningful linguistic features, train supervised classifiers, compare evaluation measures, and identify limitations and responsible-use requirements.

II. BACKGROUND AND RELATED WORK

Fake-news detection combines machine learning, NLP, data mining, information retrieval, and social-network analysis. Early systems used keyword matching and manually written linguistic rules. These systems were simple but often failed when the vocabulary, topic, or writing style changed.

Statistical classifiers improved the process by learning patterns from labeled data. Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest are commonly used for text classification. Bag-of-Words represents a document by its words, while TF-IDF gives greater importance to words that are frequent in one document but

uncommon in the full collection. N-grams preserve limited word-order information.

Recent approaches use word embeddings, neural networks, and transformer models to capture semantic and contextual relationships. A detection system should distinguish fake news from satire, opinion, parody, and ordinary reporting errors.

III. PROPOSED METHODOLOGY

A. Dataset Collection

The dataset should contain news articles with dependable labels such as Real and Fake. A record may include title, article body, author, date, source, topic, and class label. Duplicate articles, missing values, extremely short texts, inconsistent labels, and class imbalance should be examined. Near-duplicate documents must be removed to reduce data leakage.

B. Text Preprocessing

The preprocessing stage converts text to lowercase, removes irrelevant markup, tokenizes the document, and optionally removes stop words. Stemming or lemmatization can reduce related words to a common form. Title and article body should be combined consistently. Excessive cleaning should be avoided because negation words and named entities may carry useful information.

C. Feature Extraction

TF-IDF is the primary feature representation. For term t in document d, TF-IDF(t,d) = TF(t,d) x IDF(t), where IDF(t) reduces the weight of terms appearing in many documents. Unigrams, bigrams, and trigrams can capture individual words and short phrases. Additional features may include article length, sentence length, capitalized words, vocabulary diversity, readability, and sentiment scores.

D. Classification Models

Naive Bayes is fast for sparse text vectors. Logistic Regression is effective for high-dimensional TF-IDF features and provides interpretable coefficients. Support Vector Machine can separate classes in a high-dimensional space.

Random Forest combines decision trees and can model nonlinear relationships. The best model should be selected using validation results rather than accuracy alone.

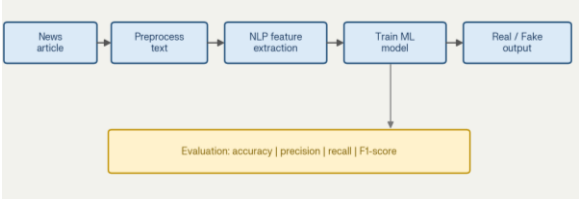

Fig. 1. Proposed fake-news detection workflow.

Table I. Qualitative comparison of candidate classifiers.

Model	Strength	Limitation
Naive Bayes	Fast for sparse text	Independence assumption
Logistic Regression	Interpretable TF-IDF weights	Linear boundary
Support Vector Machine	Strong high-dimensional separation	Needs tuning
Random Forest	Nonlinear decision rules	Less efficient for sparse text

IV. EXPERIMENTAL EVALUATION

The dataset can be divided into training and testing subsets, for example using an 80:20 split. Stratified splitting helps preserve the ratio of real and fake articles, while cross-validation can be used for hyperparameter selection. The dataset source, number of records, class distribution, preprocessing decisions, feature settings, classifier parameters, and random seed should be reported.

Accuracy measures overall correct predictions. Precision measures the proportion of articles predicted as fake that are actually fake, while recall measures the proportion of fake articles detected. The F1-score balances precision and recall. A confusion matrix shows true positives, true negatives, false positives, and false negatives. Actual values must be filled after testing the trained model.

The final results should compare classifiers and feature configurations such as unigram TF-IDF, unigram-bigram TF-IDF, and TF-IDF combined with linguistic features. Assumed accuracy values should not be included. Error analysis should examine satire, opinion, short headlines, multilingual text, and unfamiliar sources.

V. CHALLENGES AND LIMITATIONS

Dataset bias may cause a model to learn source-specific patterns instead of general properties of fake news. Fake-news creators can change vocabulary, headlines, and writing style, reducing performance on future articles. Satirical and opinion content can be incorrectly classified because it may contain exaggeration or subjective language.

Multilingual and code-switched content creates additional difficulty for models trained mainly on English. Text-only systems cannot always verify whether an event happened; an article may appear credible while making unsupported claims. Source reliability, publication date, evidence, and external knowledge may therefore be necessary.

VI. RESPONSIBLE AI PRACTICES

Fake-news detection should be used as decision support rather than as an unquestionable authority. A prediction that an article is fake means that its language resembles patterns found in previously labeled examples; it does not automatically prove that every claim is false.

Human review is important for political, medical, emergency, and legal information.

Responsible development requires transparent datasets, reliable labeling, bias testing, reporting false positives and false negatives, protection of personal information, and regular model updates. Automatic censorship should not be based only on a machine-learning score.

VII. OPEN RESEARCH DIRECTIONS

Future work can investigate multilingual and cross-lingual detection, source-credibility analysis, social-network propagation patterns, and multimodal misinformation involving text, images, audio, and video. Explainable models should identify the claim that triggered a warning and present supporting evidence.

Other directions include robustness against adversarial text manipulation, early-warning systems, standardized datasets with clear definitions of fake news and satire, and domain-specific models for health, politics, finance, and emergencies.

VIII. CONCLUSION

This paper presented a machine-learning approach for fake-news detection using NLP-based feature extraction. The proposed pipeline includes dataset preparation, text preprocessing, TF-IDF and n-gram representation, optional linguistic features, supervised classification, and evaluation using accuracy, precision, recall, F1-score, and confusion matrices.

NLP-based methods are efficient, reproducible, and comparatively interpretable. Nevertheless, performance depends on dataset quality, label accuracy, domain coverage, and changing misinformation techniques. A reliable system should combine machine learning with source analysis, evidence-based fact verification, human oversight, and responsible AI principles.

REFERENCES

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

[2] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[4] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, “Fake news or truth? Using satirical cues to detect potentially misleading news,” in Proc. 2nd Workshop Computational Approaches to Deception Detection, 2016, pp. 7–17.

[5] A. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. 2017 ACM Conf. Information and Knowledge Management (CIKM), 2017, pp. 797–806.

[6] W. Y. Wang, “Liar, liar pants on fire: A new benchmark dataset for fake news detection,” in Proc. 55th Annu. Meeting Association for Computational Linguistics (ACL), 2017, pp. 422–426.

[7] M. Granik and V. Mesyura, “Fake news detection using naive Bayes classifier,” in Proc. 2017 IEEE First Ukraine *Conf. Electrical and Computer Engineering (UKRCON)*, 2017, pp. 900–903.