International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Fake News Detection Using Machine Learning Algorithms

Libin. C[1], Muthu Arumugam. I[2], Dr. K. Sumathi[3]

[1,2]UG Student, Department of Computer Science and Data Science ,

Nehru Arts and Science College College, Coimbatore, Tamil Nadu India.

[3]Associate Professor, Department of Computer Science and Data Science,

Nehru Arts and Science College College, Coimbatore, Tamil Nadu India.

ABSTRACT

The rapid growth of social media has transformed how people access news. Although digital platforms enable fast information sharing, they also facilitate the widespread dissemination of fake news. Misinformation can distort public opinion, manipulate political processes, and erode trust in credible media. This paper proposes a binary classification framework for detecting fake and real news articles using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The system integrates static text classification, dynamic keyword-based verification, and website credibility analysis. Multiple supervised learning algorithms—including Naïve Bayes, Random Forest, Logistic Regression, and Passive Aggressive Classifier—were evaluated using benchmark datasets. Experimental results indicate that Logistic Regression achieved the best performance in static classification after hyperparameter tuning, while the Passive Aggressive classifier performed effectively for dynamic verification. The proposed system demonstrates that ML-based approaches provide a practical solution to mitigate the impact of online misinformation.

Keywords: Fake News Detection, Machine Learning, NLP, Text Classification, TF-IDF, Logistic Regression, Website Credibility

I. INTRODUCTION

The internet has become the primary source of news consumption worldwide. Social media platforms allow users to access, share, and comment on news content instantly. While this accessibility increases information reach, it also accelerates the spread of misleading and fabricated stories.

Fake news refers to intentionally false or misleading information presented as legitimate journalism. Its spread can influence elections, manipulate public perception, and generate financial profit through clickbait advertising. The 2016 U.S. presidential election highlighted the scale of this issue, where fake stories gained significant online traction.

Unlike traditional media organizations, online platforms allow low-cost content publishing with minimal verification. This creates an ecosystem where misinformation spreads rapidly through shares, bots, and coordinated campaigns.

Given these challenges, automated fake news detection has become an essential research area. Machine Learning and Artificial Intelligence have demonstrated success in text classification tasks such as sentiment analysis, spam detection, and deception detection. This study evaluates multiple ML algorithms and proposes an integrated framework combining static classification, dynamic verification, and source credibility analysis.

A. Characteristics of Fake News

Fake news articles commonly exhibit:

- Sensational or exaggerated headlines
- Emotional or provocative language
- Grammatical inconsistencies
- Unverified or unreliable sources
- Clickbait formatting

These linguistic and stylistic features make text-based machine learning classification feasible.

II. LITERATURE REVIEW

Prior research has explored multiple approaches for fake news detection:

- Granik and Mesyura applied Naïve Bayes to Facebook posts, achieving 74% accuracy but facing dataset imbalance issues.
- Gupta et al. proposed a spam detection framework using lightweight features, achieving 91.65% accuracy.
- Della Vedova et al. combined textual and social context features, reaching 81.7% accuracy in real-world deployment.
- Buntain and Golbeck analyzed Twitter threads using credibility datasets.
- Parikh and Atrey discussed psychological aspects and research challenges in fake news detection.

Existing studies confirm that text-based ML models are effective, but performance depends heavily on feature engineering, dataset quality, and model selection.

III. METHODOLOGY

The proposed system follows a three-stage architecture:

1. Static Classification Module
2. Dynamic Verification Module
3. Website Source Verification Module

A. Static Module

The static module performs supervised text classification using pre-trained ML models. The steps include:

1. Data preprocessing
2. Feature extraction
3. Model training
4. Performance evaluation
5. Hyperparameter optimization

B. Dynamic Module

This module allows users to input keywords or article text. It performs online comparison and calculates a credibility probability score using trained classifiers.

C. URL Verification

The system checks domain names against:

- Trusted source database
- Blacklisted source database

If the domain is not found, the system reports it as unrecognized.

IV. IMPLEMENTATION

A. Datasets Used

1.**LIAR Dataset**○ 12,836 annotated statements ○ Labels simplified into binary classes (True/False) ○ Train, validation, and test splits

2.**REAL_OR_FAKE Dataset**

○Used for training the Passive Aggressive classifier

○Contains labelled news statements

B. Data Preprocessing

Text preprocessing included:

- Punctuation removal
- Tokenization
- Stop word removal
- Stemming

These steps reduce noise and improve model learning efficiency.

C. Feature Engineering

Text was transformed into numerical vectors using:

1. Bag-of-Words (BoW)
2. N-Grams (Unigrams, Bigrams)
3. TF-IDF (Term Frequency–Inverse Document Frequency)

TF-IDF weighting improved discrimination between common and informative terms.

D. Classification Algorithms

The following supervised algorithms were evaluated:

1. Multinomial Naïve Bayes
2. Random Forest
3. Logistic Regression
4. Passive Aggressive Classifier (Dynamic Module)

Logistic Regression models probability using the sigmoid function:

$$logit(p) = ln\left(\frac{p}{1-p}\right)$$

Random Forest uses ensemble bagging, while Naïve Bayes applies Bayes' Theorem assuming conditional independence.

The Passive Aggressive classifier is an online learning algorithm optimized for streaming data.

V. EVALUATION METRICS

Performance was measured using:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

Definitions:

- **TP** – Correctly identified fake news
- **TN** – Correctly identified real news
- **FP** – Real news misclassified as fake
- **FN** – Fake news misclassified as real

Formulas:

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 = \frac{2PR}{P + R}$$
$$Accuracy = \frac{TP + TN}{Total}$$

3-fold cross-validation was used to ensure model robustness.

VI. RESULTS

A. Static Module Performance

Classifier	Precision	Recall	F1-Core	Accuracy
Naïve Bayas	0.59	0.92	0.72	0.60
Random Forest	0.62	0.71	0.67	0.59
Logistic Regression	0.69	0.83	0.75	0.65

After Grid Search optimization, Logistic Regression accuracy improved to approximately **80%**, making it the best static classifier.

B. Dynamic Module Performance

Passive Aggressive Classifier results:

- Accuracy: **92.73%**
- Precision: 0.93
- Recall: 0.92
- F1-Score: 0.9257

The dynamic module demonstrated high efficiency for real-time verification.

VII. CONCLUSION

This study developed a machine learning-based fake news detection system integrating static classification, dynamic keyword verification, and website credibility assessment.

Logistic Regression was identified as the most effective static classifier after hyperparameter tuning, achieving approximately 80% accuracy. The Passive Aggressive classifier showed superior performance in dynamic realtime verification with 92.73% accuracy.

The results confirm that NLP-based ML models provide a practical solution for mitigating misinformation. Future improvements may include:

- Larger multilingual datasets
- Deep learning models (LSTM, BERT)
- Real-time web crawling integration
- Social network propagation analysis

Overall, the proposed framework contributes toward improving digital information reliability.

VIII. REFERENCES

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

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

[3] 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.

[4] R. Mihalcea and C. Strapparava, “The lie detector: Explorations in the automatic recognition of deceptive language,” in *Proc. ACL-IJCNLP*, 2009.

[5] M. L. Della Vedova *et al.*, “Automatic online fake news detection combining content and social signals,” in *Proc. 22nd Conf. of Open Innovations Association (FRUCT)*, Jyväskylä, Finland, 2018, pp. 272–279.

[6] C. Buntain and J. Golbeck, “Automatically identifying fake news in popular Twitter threads,” in *Proc. IEEE Int. Conf. on Smart Cloud (SmartCloud)*, New York, NY, USA, 2017, pp. 208–215.

[7] S. B. Parikh and P. K. Atrey, “Media-rich fake news detection: A survey,” in *Proc. IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, USA, 2018, pp. 436–441.

[8] W. Y. Wang, “LIAR, Liar Pants on Fire: A new benchmark dataset for fake news detection,” in *Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada, 2017.

[9] H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, “A framework for real-time spam detection in Twitter,” in *Proc. 10th Int. Conf. on Communication Systems & Networks (COMSNETS)*, Bengaluru, India, 2018, pp. 380–383.

[10] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.