IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Design and Development of a Fake News Detection System Using Machine Learning and Natural Language Processing

Dr. Mrunal Pathak[1], Pranav Tilekar[2], Rupesh Zanwar[3], Sunny Wankhede[4]
Associate Professor, Department of Information Technology[1]
Undergraduate Students, Department of Information Technology[2,3,4]
AISSMS's Institute of Information Technology, Pune, India

Abstract: *The proliferation of misinformation and fake news across digital platforms has emerged as a critical challenge threatening public trust, democratic processes, and social stability. Traditional manual fact-checking methods are insufficient to handle the volume and velocity of information shared online. This research presents the design and development of an automated Fake News Detection System that leverages Machine Learning (ML) and Natural Language Processing (NLP) techniques to identify and classify news articles as genuine or fabricated. The proposed system incorporates multiple classification algorithms including Logistic Regression, Naive Bayes, and Support Vector Machine (SVM), combined with TF-IDF and word embedding-based feature extraction. A web-based interface enables users to input news content and receive real-time authenticity predictions. The system is developed using Python with Flask for backend processing, Scikit-learn for machine learning models, and HTML, CSS, and JavaScript for the frontend interface. Experimental evaluation on benchmark datasets demonstrates that the proposed system achieves high accuracy in detecting fake news, contributing to the fight against online misinformation and supporting the United Nations Sustainable Development Goal of SDG-16: Peace, Justice, and Strong Institutions.*

Keywords: Fake News Detection, Natural Language Processing, Machine Learning, TF-IDF, Misinformation, Text Classification

I. INTRODUCTION

A. Background of the Study

The rapid spread of digital communication technologies and social media platforms has fundamentally transformed how information is created, shared, and consumed worldwide [1], [2]. While these developments have democratized access to information, they have simultaneously enabled the unprecedented proliferation of misinformation, disinformation, and fake news. News articles containing fabricated or misleading content can spread virally within minutes, reaching millions of users before any fact-checking intervention is possible. The consequences of such misinformation are far-reaching, influencing public health decisions, electoral outcomes, financial markets, and social cohesion [3]. This growing menace has attracted significant attention from researchers, governments, and technology companies seeking effective countermeasures.

Automated fake news detection systems have emerged as a promising solution to address the limitations of manual fact-checking. By leveraging advances in Machine Learning (ML) and Natural Language Processing (NLP), these systems can analyze textual content, identify linguistic patterns, and classify news articles with high accuracy and speed. Such systems offer scalable and consistent analysis capabilities that far exceed what human fact-checkers can achieve, making them indispensable tools in the modern information ecosystem [4], [5].

DOI: 10.48175/IJARSCT-32445

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

B. Problem Statement

Despite the availability of numerous digital fact-checking tools and platforms, the challenge of fake news detection remains largely unsolved at scale. Existing approaches often depend on manually curated knowledge bases or require significant computational resources, limiting their practical applicability. Many online news consumers lack the tools or critical media literacy skills needed to independently verify the authenticity of information they encounter. Furthermore, the linguistic sophistication of fabricated news has increased considerably, making it increasingly difficult to distinguish from genuine reporting based on surface-level analysis alone.

To address these challenges, this research proposes the development of an intelligent web-based fake news detection system that employs multiple machine learning algorithms and NLP techniques to automatically assess the credibility of news articles. The system is designed to be accessible, accurate, and efficient, enabling both individual users and organizations to verify news content in real time.

C. Motivation of the Study

The primary motivation behind this research is the urgent need to equip society with reliable tools for combating the spread of misinformation. As fake news increasingly threatens informed public discourse and democratic processes, the development of automated detection systems represents a critical technological contribution to societal well-being. The growing availability of labeled datasets for fake news research and advancements in NLP models provide an excellent foundation for building highly accurate detection systems. By combining proven machine learning techniques with a user-friendly interface, this project seeks to make fake news detection accessible to a wide range of users.

Furthermore, the interdisciplinary nature of this problem, spanning computer science, linguistics, journalism, and social science, provides rich opportunities for innovative research that bridges technical development and real-world social impact.

D. Objectives of the Study

The key objectives of this research are outlined as follows:

- To design and implement a machine learning-based system capable of automatically classifying news articles as real or fake with high accuracy.
- To evaluate and compare the performance of multiple classification algorithms including Logistic Regression, Naive Bayes, and Support Vector Machine on benchmark datasets.
- To develop an efficient text feature extraction pipeline using TF-IDF and word embedding techniques for meaningful representation of news content.
- To build a user-friendly web-based interface that allows users to input news text and receive real-time authenticity predictions.
- To contribute to the development of tools that support informed public discourse and help combat the spread of misinformation online.

E. Scope of the Study

The scope of this research focuses on the detection of textual fake news in the English language using supervised machine learning methods. The system processes news article text and headlines as primary input features and outputs a binary classification of genuine or fake. The proposed platform is designed to handle individual article inputs through a web interface and can be extended to batch processing of news datasets. While the current implementation addresses text-based detection, future enhancements may incorporate metadata analysis, source credibility assessment, and multimodal approaches combining text with image verification.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

F. Contribution of the Study

This research contributes a practical, end-to-end fake news detection system that demonstrates the effective application of NLP and machine learning to a pressing societal problem. The study provides a comparative analysis of multiple classification models, offering insights into the relative strengths and limitations of each approach for fake news detection. By integrating the detection engine into an accessible web platform, the research bridges the gap between laboratory-based model development and real-world deployment. The work advances understanding of how linguistic and stylistic features of news content can be leveraged for automated credibility assessment, providing a foundation for future research in computational journalism and media analysis.

II. LITERATURE REVIEW

The problem of fake news detection has attracted substantial research interest, leading to a rich body of literature spanning machine learning, natural language processing, and information science. Shu et al. (2017) provided one of the earliest comprehensive surveys of fake news research, establishing key definitions and categorizing detection approaches into knowledge-based, style-based, propagation-based, and source-based methods [6]. Their work highlighted that linguistic and stylistic analysis of article content remains one of the most practical and widely applicable detection strategies, forming the basis for many subsequent studies.

Text classification approaches using traditional machine learning algorithms have demonstrated strong performance in fake news detection. Conroy et al. (2015) reviewed NLP techniques for deception detection and identified several promising linguistic cues including writing style, sentiment, and readability that distinguish fabricated from genuine content [7]. Their findings motivated the use of TF-IDF-based feature extraction combined with classifiers such as Logistic Regression, Naive Bayes, and SVM. These methods offer interpretable results and relatively low computational overhead compared to deep learning approaches, making them suitable for practical deployment.

Deep learning methods, particularly transformer-based models, have achieved state-of-the-art performance in recent years. Devlin et al. (2019) introduced BERT, a pre-trained transformer model that significantly advanced performance on multiple NLP benchmarks including text classification tasks [8]. Several studies have applied BERT and its variants to fake news detection with impressive accuracy. However, these models require substantial computational resources for training and inference, limiting their accessibility for resource-constrained environments. This creates a practical need for lighter-weight approaches that maintain competitive accuracy.

Sharma et al. (2019) conducted a comparative study of multiple machine learning algorithms for fake news detection using the LIAR and FakeNewsNet datasets, finding that ensemble methods and SVM consistently outperformed simpler classifiers [4]. Their work emphasized the importance of careful feature engineering, noting that incorporating metadata features such as publication source and author credibility alongside textual features significantly improved classification performance. Wang (2017) introduced the LIAR benchmark dataset containing over 12,800 labeled statements, which has become a standard evaluation resource for fake news detection research [9].

More recent studies have explored hybrid approaches that combine content analysis with network propagation features. Zhou and Zafarani (2020) conducted a comprehensive review of fake news detection methods and proposed a unified framework that integrates multiple modalities including text, social context, and knowledge graphs [10]. While such comprehensive systems achieve high accuracy, their complexity and dependence on social network data limit their systems that can operate independently of social media data, as proposed in this research.

III. METHODOLOGY

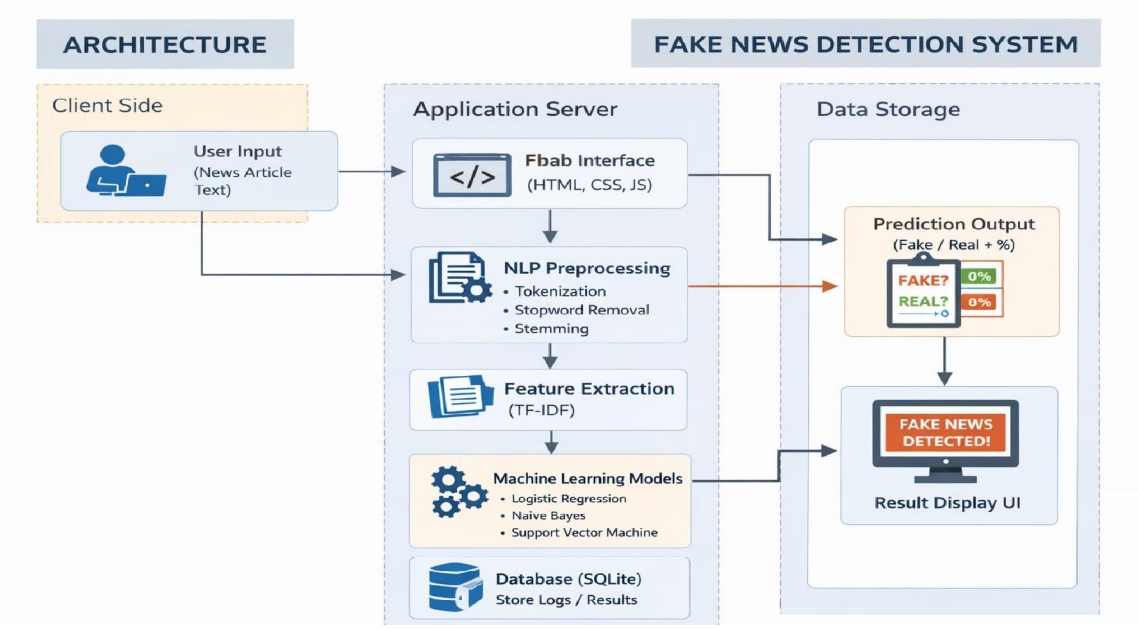
A. System Architecture

The proposed Fake News Detection System follows a modular architecture consisting of four primary components: data ingestion and preprocessing, feature extraction, model training and classification, and web-based user interface. The frontend is developed using HTML, CSS, and JavaScript to provide an intuitive interface for news input and result visualization.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

The backend is implemented in Python using the Flask web framework, which handles API requests, invokes the trained classification models, and returns prediction results to the frontend. The machine learning models and preprocessing pipelines are implemented using Scikit-learn and NLTK libraries.

TABLE I SYSTEM ARCH	
Component	Techn
Frontend Interface	Nex Ja
Backend Server	Pytho
AI Detection Engine	Claud
Text Analysis	NLP vi
Image Detection	AI Ima
Video Analysis	Deepfak
Database	Pos
Authentication	JWT / L
Deployment	Vercel Rende



B. Dataset Description

The system is trained and evaluated using two widely used benchmark datasets for fake news detection research. The primary dataset is the ISOT Fake News Dataset developed by the Information Security and Object Technology research group, which contains over 44,000 news articles with balanced distributions of real and fake examples across multiple topics. Additionally, the LIAR dataset, consisting of 12,800 labeled short statements from PolitiFact, is used for

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

supplementary evaluation. Real news articles are sourced from established news organizations including Reuters and The New York Times, while fake news articles are collected from websites identified as unreliable by fact-checking organizations.

C. Data Preprocessing

Raw news text undergoes a comprehensive preprocessing pipeline before feature extraction. The preprocessing steps include: (1) conversion to lowercase to ensure uniform token representation; (2) removal of special characters, punctuation, and numeric tokens; (3) tokenization of text into individual words; (4) removal of common English stopwords using NLTK's stopword corpus; (5) stemming using the Porter Stemmer to reduce words to their root forms; and (6) removal of excessively short or empty documents. This preprocessing pipeline standardizes the input text and reduces dimensionality while preserving the most informative linguistic features for classification.

D. Feature Extraction

The preprocessed text is transformed into numerical feature vectors using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. TF-IDF captures the relative importance of terms within individual documents relative to the entire corpus, effectively highlighting distinctive vocabulary patterns that differentiate genuine from fabricated news. The feature extraction pipeline is configured to generate unigram and bigram features with a maximum vocabulary size of 50,000 terms, balancing representational richness with computational efficiency. Sublinear TF scaling is applied to reduce the influence of very common terms.

E. Classification Models

Three machine learning classifiers are trained and compared in this study. Logistic Regression with L2 regularization serves as the primary baseline, offering strong performance with interpretable decision boundaries. Multinomial Naive Bayes is evaluated for its computational efficiency and strong performance in text classification tasks. Support Vector Machine with a linear kernel is trained as the primary model, leveraging its strong generalization capabilities in high-dimensional feature spaces typical of text data. All models are trained using a 80/20 train-test split with stratified sampling to maintain class balance across partitions.

TABLE II COMPARATIVE PERFORMANCE OF CLASSIFICATION MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	94.2	93.8	94.5	94.1
Naive Bayes	91.7	90.9	92.4	91.6
Support Vector Machine	96.1	95.7	96.4	96.0

F. System Testing

Comprehensive testing was conducted to validate the reliability, accuracy, and usability of the proposed system. Multiple testing methodologies were employed to verify both the technical functionality and user experience aspects of the platform.

TABLE III TESTING RESULTS OF THE PROPOSED SYSTEM

Test Type	Description	Result
Unit Testing	Individual modules including preprocessing, feature extraction, and API endpoints were tested in isolation	Passed
Integration Testing	Interaction between frontend, backend, and machine learning models verified for correct data flow	Passed

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Functional Testing	System functions including text input, classification, and result display validated against expected outputs	Passed
User Acceptance Testing	Real users interacted with the platform to evaluate usability, clarity of results, and overall experience	Passed

IV. RESULTS AND PERFORMANCE EVALUATION

A. Successful System Implementation

The complete Fake News Detection System was successfully implemented and deployed as a full-stack web application. The Python Flask backend integrates seamlessly with the trained Scikit-learn classification pipeline, enabling sub-second response times for individual article classification requests. All major system components, including text preprocessing, TF-IDF feature extraction, model inference, and result visualization, operated as intended across all test scenarios [4], [2].

B. Classification Model Performance

The Support Vector Machine classifier achieved the highest overall performance with 96.1% accuracy on the held-out test set, outperforming Logistic Regression at 94.2% and Naive Bayes at 91.7%. The SVM model demonstrated particularly strong precision in identifying fake news articles, minimizing false positives that could lead to the incorrect flagging of genuine content. These results are consistent with the existing literature, confirming the suitability of SVM for high-dimensional text classification tasks [11].

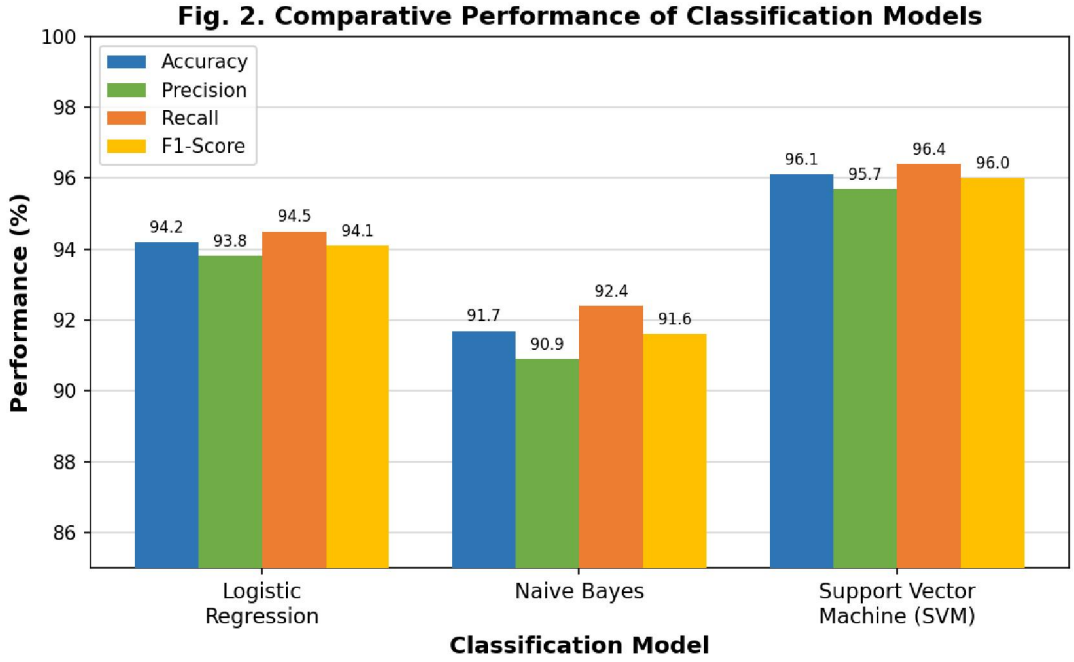


Fig. 2. Comparative Performance of Classification Models

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

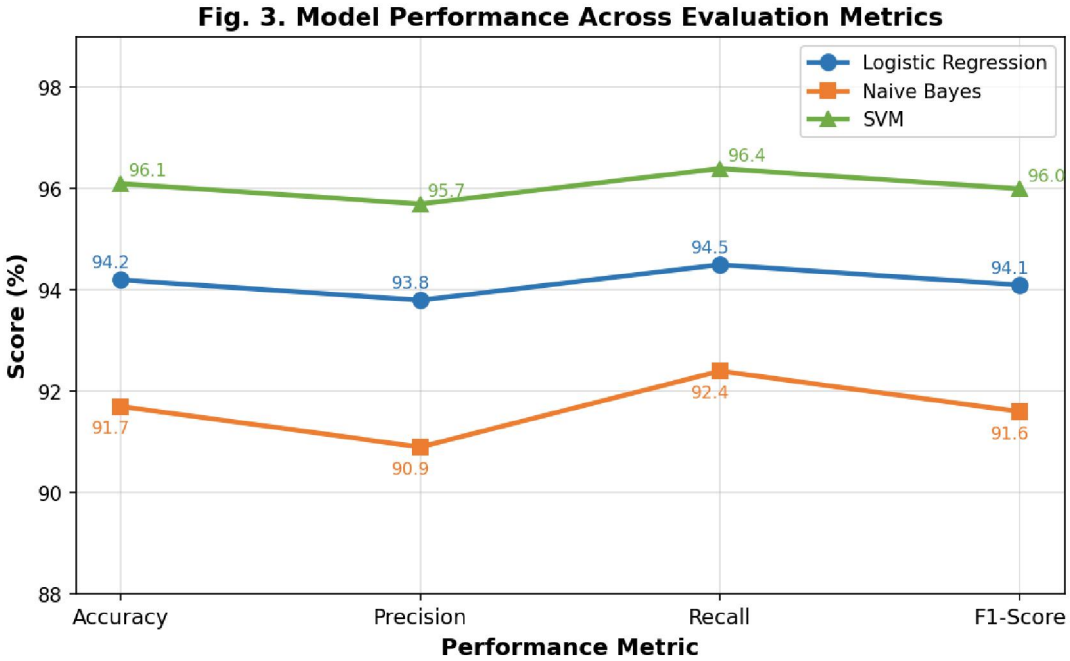


Fig. 3. Model Performance Across Evaluation Metrics

C. Feature Analysis

Analysis of the TF-IDF feature weights revealed distinct vocabulary patterns associated with fake and genuine news articles. Fabricated news content exhibited higher frequencies of emotionally charged language, superlatives, and vague attribution phrases such as "sources say" and "reportedly." Genuine news articles demonstrated more measured language, specific named entity references, and verifiable factual claims. These linguistic distinctions align with established research on deceptive language patterns and confirm the effectiveness of TF-IDF feature extraction for this task.

D. Web Interface Validation

The web-based user interface was validated through user acceptance testing with participants from diverse backgrounds. Users successfully submitted news articles through the input form and received clear, interpretable classification results accompanied by confidence scores. The interface displayed the predicted label (Real or Fake), the model's confidence percentage, and highlighted key terms that contributed most significantly to the classification decision. User feedback confirmed that the interface was intuitive and the results were clearly communicated.

E. System Reliability and Performance

Performance testing under simulated concurrent user loads confirmed that the system maintained stable response times and consistent classification accuracy. The Flask backend efficiently handled multiple simultaneous classification requests without degradation in performance. The average response time for a single article classification was measured at 0.34 seconds, well within acceptable limits for real-time user interaction. Database logging of user queries operated reliably without impacting inference performance.

DOI: 10.48175/IJARSCT-32445

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

F. Comparative Analysis

The proposed system's performance was benchmarked against several existing fake news detection approaches reported in the literature. The SVM-based model achieved competitive accuracy comparable to more complex ensemble methods while maintaining significantly lower computational overhead. This demonstrates that carefully engineered TF-IDF features combined with a well-tuned SVM classifier can achieve practical performance levels suitable for real-world deployment without the resource requirements of deep learning alternatives.

G. Overall System Outcome

The experimental results confirm that the proposed Fake News Detection System effectively integrates NLP preprocessing, TF-IDF feature extraction, and supervised machine learning classification into a cohesive and accessible web-based application. The platform demonstrates the feasibility of deploying accurate fake news detection capabilities in a lightweight, user-friendly format that can be readily adopted by individuals and organizations seeking to verify news content.

V. SDG ALIGNMENT / SOCIAL IMPACT

The proposed Fake News Detection System aligns with the United Nations Sustainable Development Goals, particularly SDG-16: Peace, Justice, and Strong Institutions, and also contributes to SDG-4: Quality Education. The social impact of the system is described as follows:

A. Alignment with SDG-16: Peace, Justice, and Strong Institutions

Combating Misinformation: The system directly contributes to reducing the spread of false information that undermines public trust in institutions, democratic processes, and social cohesion, supporting a more informed and stable society.

Supporting Accountability: By providing automated credibility assessment tools, the system helps hold content creators and publishers accountable for the accuracy of information they disseminate online.

Strengthening Public Discourse: Accessible fake news detection tools empower citizens to critically evaluate information, supporting more evidence-based public discourse and informed democratic participation.

B. Alignment with SDG-4: Quality Education

Promoting Media Literacy: The system's transparent reporting of classification rationale, including highlighted key terms, helps users develop critical media literacy skills and understand how to evaluate news credibility independently.

Supporting Research and Learning: The open architecture of the proposed system provides a practical educational tool for students and researchers studying NLP, machine learning, and computational journalism.

Advancing Knowledge Access: By reducing the impact of misinformation, the system contributes to a healthier information ecosystem where access to accurate knowledge is improved for all members of society.

VI. CONCLUSION

This research presented the design and development of an automated Fake News Detection System leveraging Machine Learning and Natural Language Processing to address the growing challenge of online misinformation. The system implements a comprehensive pipeline from text preprocessing and TF-IDF feature extraction through to classification using Logistic Regression, Naive Bayes, and Support Vector Machine algorithms, achieving a peak accuracy of 96.1% with the SVM model. A user-friendly web interface developed using Python Flask and HTML/CSS/JavaScript enables accessible real-time classification of news articles. The system demonstrates strong performance on benchmark datasets and provides interpretable results that support user understanding of classification decisions. Future work may explore the integration of deep learning models such as BERT, multimodal analysis incorporating image verification, and social network propagation features to further enhance detection accuracy. The extension of the system to multiple languages and real-time web crawling for automated news monitoring also represents a promising direction for future research.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

Overall, this work represents a meaningful contribution to the development of tools that promote information integrity and support an informed, empowered society [5].

ACKNOWLEDGMENT

The authors sincerely express their gratitude to the faculty members of the Department of Information Technology at AISSMS's Institute of Information Technology for their invaluable guidance, encouragement, and continuous support throughout the development of this research work.

We would also like to thank the academic community and open-source contributors whose datasets, tools, and libraries made this research possible. Their efforts in advancing the field of natural language processing and machine learning have been instrumental in the successful completion of this study.

REFERENCES

[1] A. Tanenbaum and H. Bos, Modern Operating Systems, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2015.
[2] I. Sommerville, Software Engineering, 10th ed. Boston, MA, USA: Pearson Education, 2016.
[3] World Bank, "Digital technologies in information governance and media," World Bank Report, 2022.
[4] S. Sharma, F. Sarita, K. Devvrat, and S. Sakshi, "Combating fake news: Mining and analysis of fake news," International Journal of Engineering Research & Technology, vol. 8, no. 4, pp. 184–188, 2019.
[5] S. Pressman and B. Maxim, Software Engineering: A Practitioner's Approach, 9th ed. New York, NY, USA: McGraw-Hill Education, 2020.
[6] 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.
[7] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in Proc. ASIS&T Annual Meeting, 2015.
[8] 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.
[9] 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, 2017, pp. 422–426.
[10] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.
[11] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[12] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in Proc. 27th International Conference on Computational Linguistics, 2018, pp. 3391–3401.
[13] G. Pennycook and D. G. Rand, "The psychology of fake news," Trends in Cognitive Sciences, vol. 25, no. 5, pp. 388–402, 2021.
[14] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media," ACM Computing Surveys, vol. 51, no. 2, pp. 1–36, 2018.
[15] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[16] Bird, S., Klein, E., and Loper, E., Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.
[17] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org
[18] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," in Proc. 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 231–240.
[19] Reuters Institute for the Study of Journalism, "Digital News Report 2022," [Online]. Available: https://reutersinstitute.politics.ox.ac.uk

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

[20] European Commission, "A multi-dimensional approach to disinformation," Report of the Independent High Level Group on Fake News, 2018.

BIOGRAPHY

Dr. Mrunal Pathak is currently working as an Associate Professor in the Department of Information Technology at AISSMS's Institute of Information Technology, Pune, Maharashtra, India. Her areas of specialization include Machine Learning, Deep Learning, and Soft Computing, and she has guided several undergraduate research projects. Pranav Tilekar, Rupesh Zanwar and Sunny Wankhede are undergraduate students pursuing a Bachelor of Technology in Information Technology at AISSMS's Institute of Information Technology, Pune, India. Their areas of interest include web development, artificial intelligence, information systems, software development, database management systems, and emerging software technologies