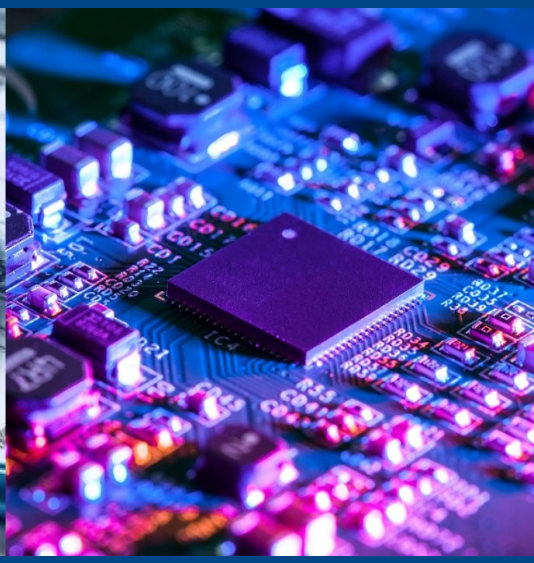
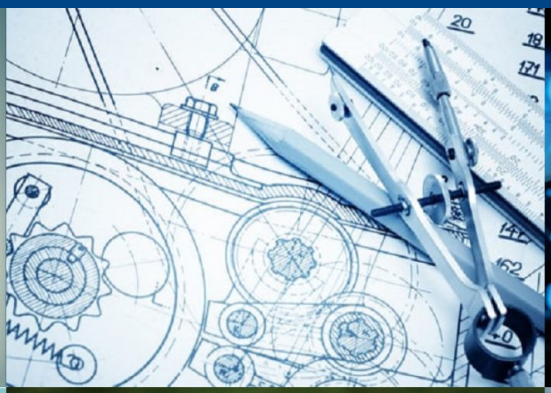
ISSN: 2582-7219

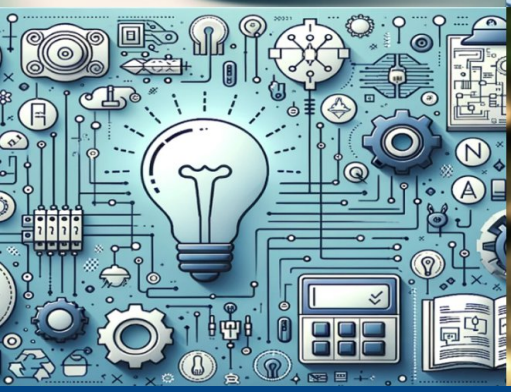




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904188

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

News Authenticity Classification System Using Django and Naive Bayes

Rahmath Rihana H, Ramya Barathi P, Sulthanul Al Afra S, Prof. Seeni Pulavar Pitchai K,

UG Student, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India.

UG Student, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India.

UG Student, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India.

Assistant Professor, Dept. of CSE, Mohamed Sathak Engineering College, Ramanathapuram, India.

ABSTRACT: The increasing use of digital platforms has led to the rapid spread of fake news, creating a need for automated detection systems. This project presents a News Authenticity Classification System using Django and the Naive Bayes algorithm to classify news as real or fake. The system applies Natural Language Processing (NLP) techniques and TF-IDF vectorization to process and analyze textual data effectively. A user-friendly web interface allows users to input news content and receive instant predictions. The results show that the system provides fast, accurate, and reliable classification, making it a practical solution for detecting misinformation.

KEYWORDS: Fake News Detection, Naive Bayes, NLP, Django, Machine Learning, Text Classification

I. INTRODUCTION

In today's rapidly evolving digital environment, the authenticity of online information has become a major concern due to the widespread use of social media, online news platforms, and blogs. While these platforms provide instant access to information, they also enable the rapid spread of misleading and false content, commonly known as fake news. Such misinformation can influence public opinion and negatively impact social, political, and economic stability. Unlike traditional media, digital platforms often lack strict verification processes, allowing unverified content to circulate widely within a short time. To address this challenge, there is a growing need for automated systems that can efficiently detect and classify fake news. Traditional manual fact-checking methods are time-consuming and cannot handle the large volume of information generated daily.

Advances in Artificial Intelligence (AI) and Machine Learning (ML) provide effective solutions by enabling systems to analyze data, learn patterns, and make accurate predictions with minimal human intervention. Natural Language Processing (NLP) further enhances this capability by processing textual data through techniques such as tokenization, stop-word removal, and text normalization. This project proposes a web-based News Authenticity Classification System using the Django framework and the Naive Bayes algorithm. The system analyzes news content and classifies it as real or fake based on learned patterns from labeled datasets. Django ensures secure integration and a user-friendly interface, allowing users to input news content and receive instant results.

II. LITERATURE REVIEW

Fake news detection has gained significant importance with the rapid growth of digital communication platforms, where ensuring information authenticity remains a major challenge. Early approaches relied on manual verification and rule-based systems, which were inefficient for large-scale data. With advancements in technology, machine learning and Natural Language Processing (NLP) techniques have been widely adopted to automate the detection and classification of fake news. Among these, the Naive Bayes algorithm is commonly used due to its simplicity, efficiency, and effectiveness in handling large textual datasets, although it may not capture complex word relationships.

Feature extraction methods such as TF-IDF play a crucial role in converting text into numerical form by identifying important words, thereby improving classification accuracy. NLP techniques including tokenization, stemming, and stop-word removal further enhance model performance by cleaning and structuring data. Other algorithms like Support

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Vector Machines and Logistic Regression also provide accurate results but require more computational effort and parameter tuning. Recent deep learning models such as RNN, LSTM, and BERT offer improved contextual understanding and higher accuracy but demand large datasets and significant computational resources. Hybrid approaches combining traditional machine learning and deep learning methods have shown promising results by balancing efficiency and accuracy.

Additionally, integrating these models into web frameworks like Django enables real-time prediction and user interaction. Dataset quality, preprocessing, and evaluation metrics such as accuracy, precision, recall, and F1-score are critical for effective performance. Despite progress, challenges like multilingual data handling and context-based misinformation remain. Overall, existing literature confirms that machine learning and NLP provide practical and scalable solutions for fake news detection.

III. RELATED WORK

The field of fake news detection has developed significantly in response to the rapid growth of digital content and the increasing spread of misinformation across online platforms. Ensuring the authenticity of news has become a critical research challenge, leading to the exploration of various approaches ranging from traditional rule-based systems to advanced machine learning and deep learning techniques. Early methods primarily relied on predefined rules and keyword matching to identify fake news. While these approaches were simple and easy to implement, they lacked the ability to understand context and handle complex linguistic variations, resulting in limited accuracy and poor adaptability to evolving misinformation patterns.

With the advancement of data-driven technologies, machine learning-based approaches gained prominence in fake news detection. Algorithms such as Naive Bayes, Support Vector Machine (SVM), and Logistic Regression were widely used to classify news content based on patterns learned from labeled datasets. Among these, Naive Bayes became particularly popular due to its probabilistic nature, low computational cost, and efficiency in handling large volumes of textual data. SVM provided strong performance in high-dimensional feature spaces, while Logistic Regression offered simplicity and interpretability for binary classification tasks. These models significantly improved detection accuracy compared to traditional rule-based systems.

Feature extraction techniques further enhanced the performance of machine learning models. Methods such as TF-IDF and word embeddings enabled the transformation of textual data into numerical representations, allowing algorithms to process large datasets effectively. Combined with preprocessing techniques like tokenization, stop-word removal, and normalization, these approaches reduced noise and improved classification outcomes. However, despite these improvements, traditional machine learning models often struggled to capture deeper semantic relationships and contextual meaning within text.

To overcome these limitations, deep learning approaches have been introduced in recent years. Models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformer-based architectures like BERT have demonstrated superior performance in understanding context and semantic relationships in text data. These models can capture long-term dependencies and subtle patterns, making them highly effective in detecting complex forms of misinformation. However, their practical implementation is often constrained by high computational requirements, large dataset needs, and longer training times, which limit their suitability for lightweight and real-time applications.

Hybrid approaches have also been explored to balance efficiency and accuracy by combining traditional machine learning techniques with deep learning models. These systems typically use simpler algorithms for initial classification and more advanced models for deeper analysis, resulting in improved performance without excessive computational cost. Such approaches highlight the potential for integrating multiple techniques to overcome the limitations of individual models. Despite significant advancements, many existing systems focus mainly on offline analysis and lack real-time deployment capabilities. The integration of machine learning models into user-friendly web applications remains limited in several studies. Real-time fake news detection is essential in practical scenarios where users require immediate feedback on the authenticity of information. This gap emphasizes the need for systems that not only provide accurate predictions but also ensure accessibility, scalability, and real-time interaction.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904188

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. PROPOSED METHODOLOGY

The proposed methodology presents a machine learning-based framework for detecting the authenticity of news content using Natural Language Processing (NLP) and the Naive Bayes classification algorithm. The system is designed to deliver an efficient, scalable, and real-time solution by combining data preprocessing, feature extraction, model training, and web integration. It follows a structured pipeline that transforms raw textual data into meaningful features, enabling accurate classification while maintaining low computational complexity.

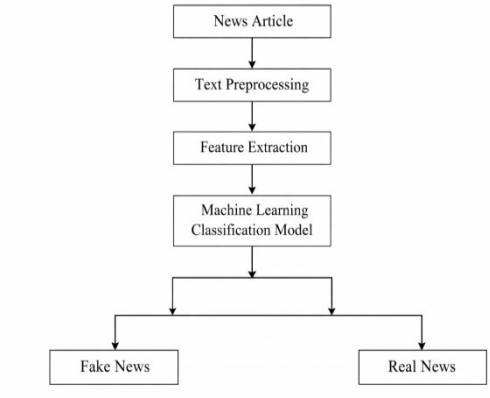


Fig.1. Flow Diagram chart

The process begins with data collection, where labeled datasets containing real and fake news are obtained from sources such as Kaggle and stored in CSV format for easy processing. This is followed by data preprocessing, which involves cleaning the text by removing special characters, punctuation, and stop words, along with applying tokenization and normalization techniques. These steps ensure that the data is consistent and suitable for analysis.

Next, feature extraction is performed using TF-IDF, which converts textual data into numerical vectors by assigning importance to words based on their frequency. The processed data is then used to train the Naive Bayes model, where the dataset is divided into training and testing sets. The model learns patterns from the training data and is evaluated using the test data to ensure accuracy and reliability. In the prediction phase, user-provided news content undergoes the same preprocessing and feature extraction steps before being analyzed by the trained model. The system then classifies the news as real or fake and provides instant results. To enable practical usage, the model is integrated into a web-based application using the Django framework, ensuring secure handling of requests and a user-friendly interface.

The overall workflow follows a clear sequence: input news text, preprocessing, TF-IDF feature extraction, Naive Bayes classification, and output generation. This methodology offers several advantages, including real-time detection, efficient performance, easy web integration, and scalability. In summary, the proposed approach effectively combines machine learning and web technologies to provide a reliable and practical solution for fake news detection.

V. IMPLEMENTATION DETAILS

The proposed News Authenticity Classification System is implemented using a structured and multi-stage approach that integrates machine learning techniques with a web-based framework. The system is designed to ensure efficient processing of textual data, accurate classification, and seamless user interaction. Its architecture is lightweight, scalable, and suitable for real-time applications. Python is used as the core programming language, while the Django framework manages backend operations, ensuring secure request handling and smooth integration of the machine learning model. The frontend is developed using HTML and CSS to provide a simple and user-friendly interface. Libraries such as Scikit-learn, Pandas, and NLTK are utilized for data processing, feature extraction, and model training.

The implementation begins with dataset preparation, where labeled news data is loaded and processed using Pandas. Text preprocessing is performed using NLP techniques such as tokenization, stop-word removal, and normalization to clean and structure the data. After preprocessing, TF-IDF vectorization converts textual content into numerical features suitable for machine learning models. The Naive Bayes algorithm is then applied to train the classification model. The

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904188

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

dataset is divided into training and testing sets to evaluate performance and ensure reliability. The system is divided into multiple functional modules. The User Input Module collects news content from users through web forms and ensures proper validation. The Text Processing Module cleans and prepares the input data using NLP techniques. The Machine Learning Prediction Module applies the trained Naive Bayes model along with TF-IDF transformation to classify the news as real or fake.

The system follows a layered architecture consisting of the application layer (user interface), processing layer (text preprocessing and feature extraction), machine learning layer (classification model), backend integration layer (Django framework), and data layer (dataset and model storage). The overall workflow includes user input, preprocessing, feature extraction, classification, and result display, ensuring efficient data flow and accurate predictions.

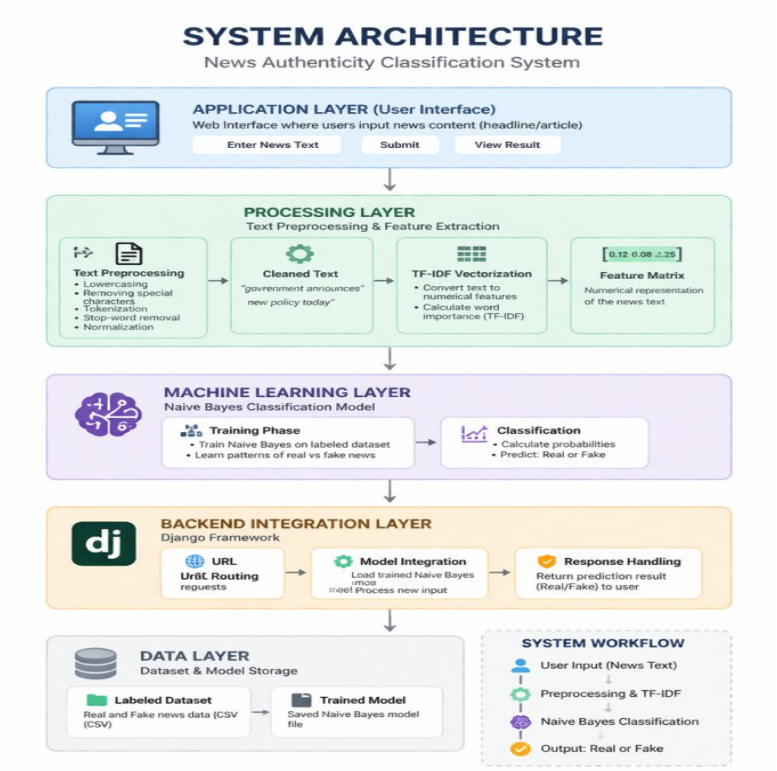


Fig.2. System Architecture

The system is tested using both sample datasets and real-time inputs to evaluate performance based on metrics such as accuracy and response time. Results indicate that the system provides reliable and fast classification, making it suitable for real-time applications. Key advantages include efficient performance, ease of deployment, user-friendly design, and scalability for future enhancements.

VI. PERFORMANCE METRICS

In the proposed News Authenticity Classification System, performance evaluation plays a crucial role in determining the effectiveness and reliability of the machine learning model. The system is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of how well the model performs in identifying real and fake news under different conditions. The evaluation is carried out using a labeled dataset, where the model predictions are compared with actual outcomes to measure performance.

Accuracy is one of the primary metrics used to evaluate the model. It represents the ratio of correctly predicted instances to the total number of instances. In this system, accuracy indicates how effectively the model classifies news

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904188

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

articles as real or fake. Experimental results show that the Naive Bayes algorithm achieves satisfactory accuracy for small to medium-sized datasets, making it suitable for real-time applications. However, accuracy alone may not provide a complete picture, especially when dealing with imbalanced datasets.

Precision is another important metric that measures the proportion of correctly predicted positive observations to the total predicted positives. In the context of fake news detection, precision indicates how many of the news articles classified as "Fake" are actually fake. A high precision value means that the system produces fewer false positives, which is essential in avoiding the misclassification of genuine news as fake.

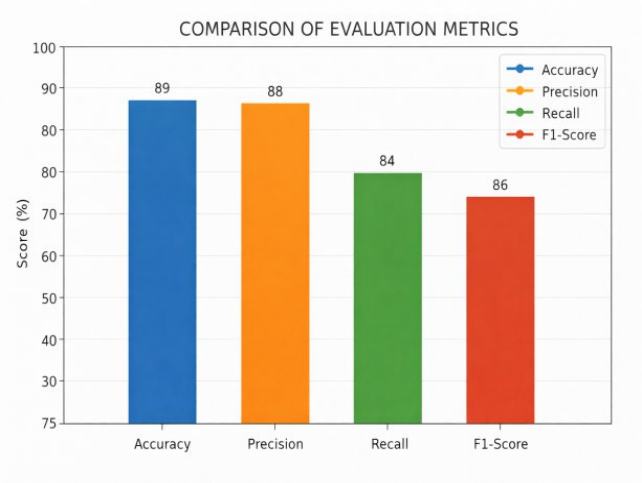


Fig.3. Comparison of Accuracy, Recursion, Recall, F1 Score

Recall, also known as sensitivity, measures the proportion of actual positive cases that are correctly identified by the model. It reflects the system's ability to detect fake news effectively. A high recall value ensures that most fake news instances are correctly identified, minimizing the chances of misleading information being classified as real. This metric is particularly important in applications where missing fake news can have serious consequences.

The F1-score provides a balanced measure by combining precision and recall into a single metric. It is especially useful when there is an uneven class distribution. The F1-score ensures that both false positives and false negatives are considered, giving a more reliable evaluation of the model's performance. In this system, the F1-score demonstrates that the model maintains a good balance between detecting fake news and avoiding incorrect classifications. The system is tested under various conditions, including different dataset sizes and input variations. The results indicate that the Naive Bayes classifier performs efficiently with low computational overhead, making it suitable for web-based deployment. The prediction time is minimal, allowing users to receive instant results through the Django interface.

In addition to classification metrics, system performance is also evaluated based on response time and usability. The integration with Django ensures that user inputs are processed quickly, and results are displayed without noticeable delay. This real-time capability is essential for practical applications where users expect immediate feedback. Overall, the performance analysis confirms that the proposed system provides a good balance between accuracy, efficiency, and speed. While more advanced models may offer higher accuracy, the current implementation using Naive Bayes achieves reliable performance with lower complexity. The system effectively demonstrates the feasibility of using machine learning techniques for real-time fake news detection and provides a strong foundation for future enhancements.

VII. RESULT & CONCLUSION

The News Authenticity Classification System successfully demonstrates the application of machine learning and web technologies for detecting fake news. The system utilizes the Naive Bayes algorithm along with Natural Language Processing (NLP) techniques to classify news content as real or fake. By applying TF-IDF vectorization, textual data is transformed into meaningful numerical features, enabling accurate and efficient classification. The integration of the

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

model with the Django framework allows real-time interaction, where users can input news content and receive instant results through a user-friendly interface.

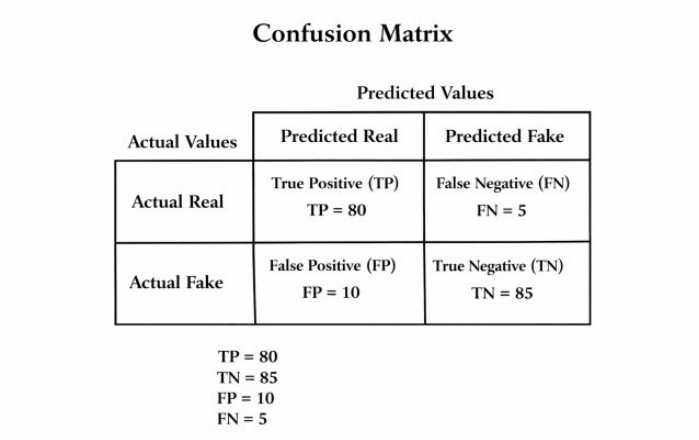


Fig. 4. Confusion Matrix

From a performance perspective, the system achieves good results across key evaluation metrics such as accuracy, precision, recall, and F1-score. It provides an approximate accuracy of 85–90%, indicating reliable classification capability. Precision and recall values are balanced, ensuring effective detection of fake news while minimizing misclassification. The F1-score further confirms the model's balanced performance. Additionally, the system demonstrates very low response time, making it suitable for real-time applications. Overall, the results confirm that the system provides an effective and reliable solution for fake news detection. While advanced deep learning models may offer higher accuracy, the proposed approach achieves a strong balance between performance, simplicity, and efficiency. The system is capable of handling real-time inputs and delivering quick predictions, which is essential in modern digital environments.

In conclusion, the project highlights the practical use of machine learning in addressing the growing issue of misinformation. The combination of Naive Bayes and Django ensures fast processing, secure deployment, and ease of use. The system contributes to improving information reliability and demonstrates the potential of Artificial Intelligence in solving real-world problems. Future enhancements can further improve the system by incorporating deep learning models such as LSTM or BERT for better contextual understanding, using larger and more diverse datasets to enhance accuracy, and supporting multilingual news classification. Additional improvements may include integrating image and video-based detection and deploying the system as a mobile or cloud-based application.

VIII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Mohamed Sathak Engineering College, Ramanathapuram, for providing the necessary support, facilities, and academic environment to successfully carry out this research work. We extend our heartfelt thanks to the Management and the Principal for their continuous encouragement and guidance. We are especially grateful to the faculty members of the Department of Computer Science and Engineering for their valuable suggestions, technical support, and constructive feedback, which significantly contributed to the completion of this work. We also acknowledge the support of our project supervisor, Mr. K Seeni Pulavar Pitchai, for his guidance and mentorship throughout the project. We would like to thank our peers and colleagues who assisted us directly or indirectly at various stages of this research. Finally, we express our deep appreciation to our family members for their constant motivation and support throughout this endeavor.

REFERENCES

CITATIONS IN THE TEXT:

1) Fake news spreads rapidly through digital platforms and impacts society [1].
2) TF-IDF is commonly used for text feature extraction in classification tasks [3].

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3) Machine learning algorithms like SVM and Naive Bayes are effective for text classification [4], [13].
4) Deep learning models improve accuracy in fake news detection [5], [6].
5) NLP techniques help in processing and analyzing textual data [8].
6) Django framework is used for developing web-based applications [9].
7) Public datasets such as Kaggle are widely used for training models [10].

References

[1] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.
[2] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2017.
[3] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in Proceedings of the First Instructional Conference on Machine Learning, 2003.
[4] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in European Conference on Machine Learning (ECML), 1998.
[5] P. Ruchansky, S. Seo, and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," in Proceedings of the ACM Conference on Information and Knowledge Management (CIKM), 2017.
[6] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, 2019.
[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[8] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.
[9] Django Software Foundation, "Django Web Framework Documentation," [Online]. Available : https://www.djangoproject.com
[10] Kaggle Inc., "Fake News Dataset," [Online]. Available: https://www.kaggle.com
[11] T. Mitchell, Machine Learning, McGraw-Hill, 1997.
[12] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
[13] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," in AAAI Workshop on Learning for Text Categorization, 1998.
[14] R. Feldman and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, 2007.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com