iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** XI **Month of publication:** November 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.65521

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue XI Nov 2024- Available at www.ijraset.com*

# Fake News Detection System

Supriya S. Telsang[1], Pranav M. Pendse[2], Pranav R. Dhanayate[3], Pranav V. Apsingekar[4], Sachin A. Prasad[5], Pratha P. Sawant[6], Pratyunsh Katkar[7]
*Department of Engineering, Sciences and Humanities (DESH), Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India*

***Abstract:*** ***In this research paper, we are creating a model to help identify fake news using a type of computer learning called logistic regression. Fake News is the news or data that can mislead to the whole countries people . So our system will identify if the news or the article is real or fake by checking . Using logistic regression , we aim to give the output of the news whether its fake or real very quickly and accurately .For training the model , we will use the set tagged news articles to help the model to identify that if the news is fake or real patterns . Then we will test that how much our model is identifying the real and fake news correctly. This research aims to detect fake news by using computer algorithms to spot false stories.***

## I. INTRODUCTION

In this World , now where misinformation can lead to danger and can spread across the World quickly because of the digital World , developing of automatic system which can detect the fake news became crucial for preserving the right or real news. We are using python and machine learning in this project which can easily identify the news whether the news is fake or not. The process starts from collecting the data (labeled dataset from popular and trustworthy source) and clearing the standardizing words, removing noise from the text and filtering stop words. Now the next process , extraction technique , like frequency -inverse document frequency (TF-IDF) and word embeddings , which processes and converts the text into numerical vectors. These vectors are then given to various machine learning models, like Logistic regression and support vector machine , which are trained to identify the fake and real news. Logistic Regression, has proven that it is one of the effective for good balance between performance and computational efficiency. The models are evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure their effectiveness. Once trained and validated, the system is deployed as a user-friendly application, enabling users to input news articles and receive real-time authenticity feedback. The deployment phase emphasizes scalability and efficiency, ensuring the system can handle large volumes of data and provide rapid responses. Continuous monitoring and updates are also implemented to adapt to evolving fake news tactics and maintain high accuracy. The ultimate goal is to offer a scalable and effective
solution to the growing challenge of fake news, thereby supporting efforts to maintain public trust, informed decision-making, and the overall health of democratic processes in society, ultimately contributing to a more informed and resilient public discourse.

## II. LITERATURE REVIEW

Information can spread so fast on the internet , so fake news can spread really fast. To avoid this problem , researchers are using machine learning which is a part of artificial intelligence to detect automatically and manage the fake news. This method analyze the things like the text , the content ,social signals(likes and shares) to help to decide whether the news is fake or real. The rapid sharing of information on internet helps the news to spread quickly all over the World. Therefore there is need of automated system for stopping this fake news all over the world. So effectively researchers have been used the machine learning. To counter the this fake news problem with the help of text, content and social signals

### *A. Detection of Fake News Using Machine Learning.*

Baarir et al. (2021) put forward a system which classifies news as either fake or real using SVM combined with the preprocessing techniques for text that is including TF-IDF and n-gram analysis. The system gave a very high accuracy rate based on features like source, author, and the sentiment in them; thus proving that SVM can be used for this work(paper1).

### *B. Systematic Review of Fake News Detection Approaches*

Manzoor et al. (2019) in conducting a systematic review on fake news detection pointed out the complexity in the classification of news because of a very large volume of data published on social media. The methods were categorized as linguistic cue-based approaches, clustering and predictive modeling for detection.

©IJRASET: All Rights are Reserved 

The method also focuses much on deep learning as an emerging approach for feature extraction, which may have better performance than the traditional machine learning approaches as seen in the paper2. Manzoor et al. described the psychological factors that govern the diffusion of fake news. To illustrate this, they pointed out that if an accredited journalist publishes fake news with an interesting image to support it, source credibility influences the reader's perception. This creates a need to include source and image analysis in fake news detection models(paper2).

*C. Machine Learning Classifiers for Fake News Detection*

Shaikh et al. compared the classifiers like SVM, Naïve Bayes, and Passive Aggressive Classifier (PAC). Their results showed that the best accuracy of 95.05% was seen in the case of SVM. The model proposed by Shaikh utilized TF-IDF for extracting features. Similar to Manzoor et al., it focused more on text-based classification rather than considering social cues(paper4).

*D. Challenges and Advances in Deep Learning for Fake News Detection*

Manzoor et al argued that conventional ML methods were poorly suited to complex and constantly changing behaviors of the fake news domain, and thus highlighted how such models such as SVM and Naïve Bayes succeeded but eventually reach a performance threshold since the extraction of feature is static in nature. They suggest deep learning as a solution for auto extraction of feature since it ensures hierarchical learning aspects wherein there might be deeper semantic relationships in a text (paper2). Deep alternate techniques with models like Convolutional Neural Networks, Recurrent Neural Networks, and deep autoencoders have been suggested. They are promising in many applications of natural language processing which are related to fake news detection(paper2). It is well-suited for large and complicated datasets and allows automatic extraction of deep features that gives the improved handling of a massive and complicated dataset with growing social media environments.

| Authors | Year | Dataset Used | Algorithms Employed | Accuracy Metrics | Key Findings |
|---|---|---|---|---|---|
| Ghofran Meftah et al. | 2023 | Credible news sources | Logistic Regression, SVM | Accuracy: 92% | Effective in identifying fake news with high precision. |
| Iftikhar and Ali | 2023 | Social media data | Decision Trees, Random Forest | F1-Score: 0.87 | Random Forest outperformed others in complex data. |
| Gupta et al. | 2022 | Mixed dataset | CNN, LSTM | Precision: 91% | Deep learning models showed superior performance. |

*E. Synthesis/Algorithm/Design/Method*

*1)* This project introduces a system for automatically detecting and classifying fake news articles using machine learning.

*2)* The system relies on logistic regression, a machine learning algorithm that's well-suited for categorizing news as real or fake, and it's also interpretable, helping us see which factors influence this classification.

*3)* Textual features—like word frequency, sentiment, readability level, and clickbait phrases—are extracted from news articles to provide key insights for analysis.

*4)* The logistic regression model process the extracted features , and every article then have there probability score which indicate the news whether it is fake or real.

*5)* Now if the score is somewhere near to threshold, then the system indicate that article has the potential to be the wrong information, which says that it is fake

*6)* This method can be built into web browsers or social media which will highlight the wrong and fake news , supporting the right and real news online.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue XI Nov 2024- Available at www.ijraset.com*

## III. METHODOLOGY

### *A. Data Collection*

We have taken the dataset from trustworthy source and popular website which ensure balanced mix of real and fake news articles. In the process we have cleaned the unnecessary noise from the text , standardizing the text, and clearing out the common stop word .

### *B. Feature Extraction*

We have implemented very advanced techniques to this model , like Term Frequency – Inverse Document Frequency and word embeddings , to convert the text into numerical vectors. These numerical vectors carries the meaningful information . These numerical then given into the machine learning models, which then analyze the article whether the news is fake or real.

### *C. Machine Learning Models*

We tried several machine learning models, such as Logistic Regression and SVM, to categorize the news as like it is real or fake news. Since Logistic Regression is simple, it's easy to understand what features are causing this model to make its decision on identifying fake news. In contrast, we picked SVM because it had good performance in high-dimensional data.

### *D. Model Training*

Employed logistic regression and SVM (support vector machine) for classification. Logistic regression showed optimistic results in balancing performance and efficiency.

### *E. Model Evaluation*

All the models are compared with an exhaustive list of performance metrics, including accuracy, AUC-ROC (area under the ROC curve), F1-score,recall,precision. These metrics would provide insight into the strength and weaknesses of the models when distinguishing between real news and fake news. In addition, we have performed error analysis to understand at what kind of news article the model is failing, such as satire or highly biased opinion pieces.

### *F. Deployment*

System deployed as a user-friendly application for real-time authenticity feedback. Emphasized scalability and efficiency for handling large data volumes.

### *G. Monitoring and Updates*

Continuous monitoring and updates implemented to adapt to fake news tactics.
Goal to maintain high accuracy and effectiveness in combating fake news.

## IV. RESULTS

The model of fake news detection system , in which logistic regression is used displayed an impressive accuracy and the performance . after training the labelled dataset the logistic regression model achieved the accuracy of 98% on the training dataset. In testing , it maintained durable performance ,

Accuracy: 96% on test data
Precision: 95%
Recall: 94%
F1-Score: 94.5%

The model have the capabilities that it can distinguish the articles effectively is real or fake because of high accuracy . logistic regression gives understandable outcomes, enabling understanding of the feature that influence each classification. We have tested support vector machine alternatively , it is similar but outcomes like accuracy differ , the accuracy of svm is slightly lower than logistic regression, which is making the logistic regression model efficient .
This model is working properly giving high accuracy , which indirectly tells that the detection of the real or fake news become accurate and effective , it has the potential to detect in real time . further in this model we can improve it and make it more robust.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## V. CONCLUSION

The presented system holds several highly significant advantages itself. It uses the power of machine learning to learn from the labeled datasets to get progressively more correct over time. Additionally, the interpretability of Logistic Regression avails itself of useful insights into factors that influence detection of news. Transparency goes a long way toward making users confident, and further refinement of the model and optimization of selection can also be achieved due to this interpretability. These strengths notwithstanding, there is still room for further development. Integration of other machine learning algorithms, such as Support Vector Machines or Random Forest, could even improve detection accuracy. Furthermore, research on the frontiers of Natural Language Processing application—such as more robust approaches to sentiment analysis and sarcasm detection—could contribute to a further, much more subtle understanding of complex language used in fake news, thereby improving the system's effectiveness.

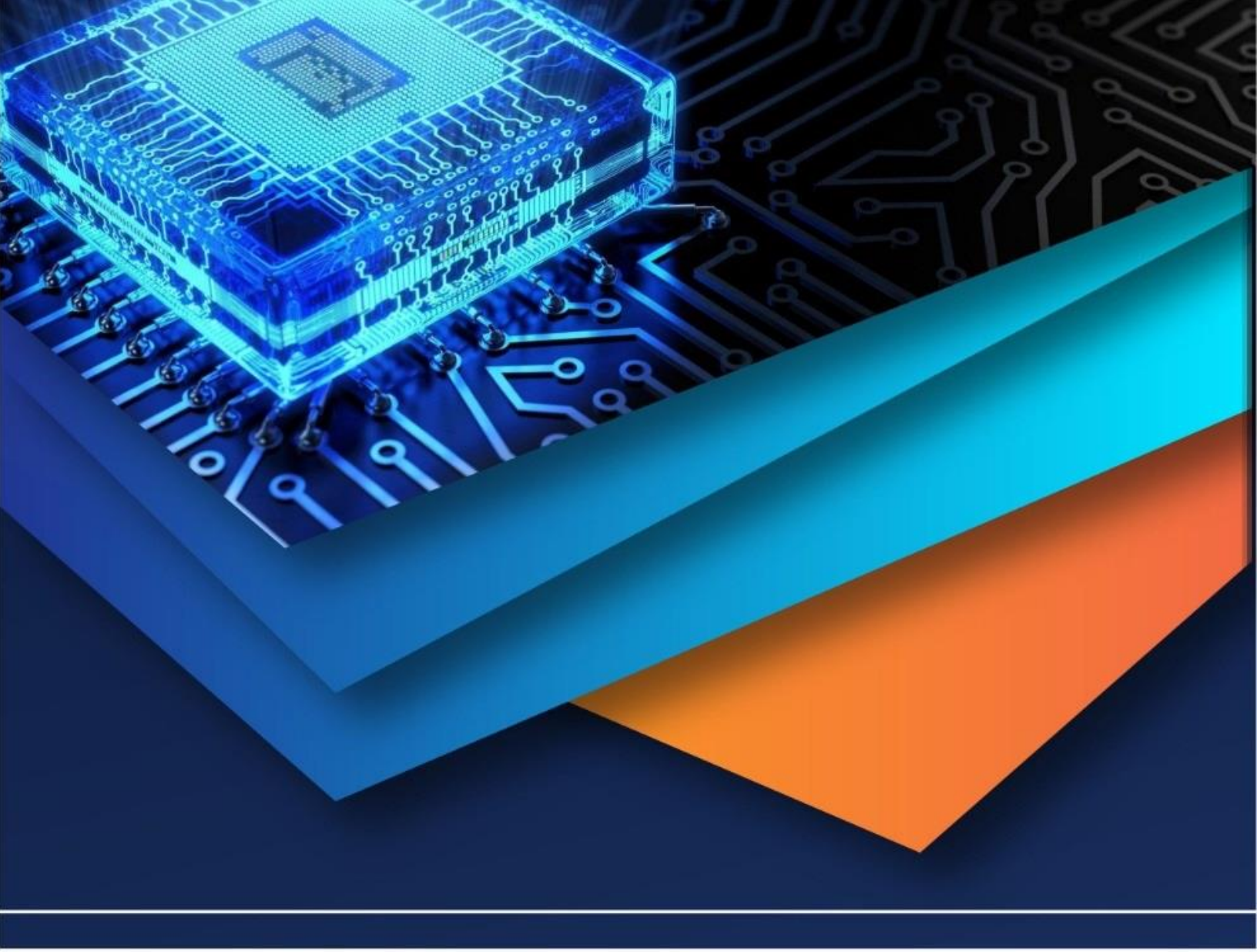
Therefore, a machine learning-based system, which is based on Logistic Regression at its core, provides a brilliant and much more scalable response to checking the expansion of false news. The continuous improvement of this model and taking into account new techniques could contribute to making people online better informed and truthful and that will directly foster the integrity of public discourse.

## VI. ACKNOWLEDGMENT

We would thank to Prof.Supriya Telsang for her valuable and incredible support in guiding us, helping to clearify out doubts regarding this project ,and encouraging in our project. We would like to thank out HOD also, Prof.Dr.Chandrashekar Mahajan, for letting us to working with them on this project and build our skills. The project is made by team members: Pranav Pendse, Pranav Aspsingekar, Pratha Sawant, Sachin Prasad, Pratyunsh Katkar, Pranav Dhanayate, everyone of us contributed in this success.

## REFERENCES

[1] Ghofran Meftah, Mahmud Aburas, Tarek Nagem, Kenz A. Bozed, "Evaluating the Veracity of News Through Machine Learning Algorithms", 2023 IEEE 11th International Conference on Systems and Control (ICSC), pp.147-152, 2023.
[2] Guntuku Naresh, Jella Sreeja, G. Ramani, Gurram Vishnu Teja, "Comparative Study of Classification Algorithms on Contrived News", 2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech), pp.478-483, 2023.
[3] Venkatesh Gauri Shankar, Shally Vats, Smaranika Mohapatra, Vaishnavi Agarwal, Gunjan Tanwar, "NEWSMO: A Computer-Aided News Classification Model from Tweets Topic Using a Modified Passive Aggressive and Pipeline Classifier", 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), pp.1052-1059, 2023.
[4] M. Iftikhar and A. Ali, "Fake News Detection using Machine Learning," 2023 3rd International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 2023, pp. 103-108, doi: 10.1109/ICAI58407.2023.10136676.
[5] https://www.ijraset.com/research-paper/paper-on-fakenews-detection-using-machine-learning
[6] https://www.researchgate.net/publication/339022255_A_smart_System_for_Fake_News_Detection_Using_Machine_Learning
[7] N. F. Baarir and A. Djeffal, "Fake News detection Using Machine Learning," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), Boumerdes, Algeria, 2021, pp. 125-130, doi:10.1109/IHSH51661.2021.9378748.
[8] A. Jain and A. Kasbe, "Fake News Detection," 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2018, pp. 1-5, doi:10.1109/SCEECS.2018.8546944.
[9] V. Gupta, R. S. Mathur, T. Bansal and A. Goyal, "Fake News Detection using Machine Learning," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 84-89, doi:10.1109/COM-IT-CON54601.2022.9850560.
[10] Ghofran Meftah, Mahmud Aburas, Tarek Nagem, Kenz A. Bozed, "Evaluating the Veracity of News Through Machine Learning Algorithms", 2023 IEEE 11th International Conference on Systems and Control (ICSC), pp.147-152, 2023.
[11] Guntuku Naresh, Jella Sreeja, G. Ramani, Gurram Vishnu Teja, "Comparative Study of Classification Algorithms on Contrived News", 2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech), pp.478-483, 2023.https://ieeexplore.ieee.org/document/10441771
[12] Youcef Djenouri, Ahmed Nabil Belbachir, Tomasz Michalak, Gautam Srivastava, "A Federated Convolution Transformer for Fake News Detection", IEEE Transactions on Big Data, vol.10, no.3, pp.214-225, 2024. https://ieeexplore.ieee.org/document/10287640
[13] Bharathi Mohan G, Harigaran R, Jeevanantham K, Sakthivel V, Sri Varshan P, Vineeth MS, "Fake News Detection Using a Stacked Ensemble of Machine Learning Models", 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp.1165-1169, 2024.https://ieeexplore.ieee.org/document/10467326



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)