

**Impact Factor:
6.007**



Date: 30-Jun-2026

International Journal of Education & Applied Sciences Research

International Peer Reviewed & Refereed Journal as per UGC New Rules

Certificate of Publication

This is to certify that the article

**“SEMANTIC ANALYSIS-BASED FAKE NEWS DETECTION USING NLP AND
MACHINE LEARNING”**

by

Dr. H. Jayamangala,
Assistant Professor, DCA PG, VISTAS, Chennai.
Email: jayamangala.scs@vistas.ac.in

has been published in ARSEAM's Journal

**“International Journal of Education & Applied Sciences
Research”, Volume 13, Issue 01, 2026.**

The Journal has been registered and assigned ISSN (Online): 2349 -2899 and
ISSN (Print): 2349-4808

Managing & Publishing Editor

**Academic Research In Science,
Engineering, Art & Management
(ARSEAM) Foundation**

ARSEAM foundation aims to publish research paper, article, case study and book review of very high quality intended to serve as an outlet for theoretical and empirical research in all areas of Science, Engineering, Art and Management contributing to the advancement in the existing literature and are very significant in managerial implications.

For detail, please visit us at: www.arseam.com



International Journal of Education & Applied Sciences Research

ISSN: 2349 -2899 (Online) ISSN: 2349 - 4808 (Print)

An International Open Access, Peer Reviewed, Refereed Journal

Volume-13 | Issue-1 | 2026|

DOI : <https://doi.org/10.5281/zenodo.22268128>

Journal homepage: <https://arseam.com/journal?title=IJEASR>

SEMANTIC ANALYSIS-BASED FAKE NEWS DETECTION USING NLP AND MACHINE LEARNING

Dr. H. Jayamangala,

Assistant Professor, DCA PG, VISTAS, Chennai.

Email: jayamangala.scs@vistas.ac.in

Received: 12 Apr 2026, **Revised:** 22-May-2026, **Accepted:** 19 Jun 2026, **Published online:** 30-Jun-2026

Cite this Article as: H. Jayamangala (2026), “Semantic Analysis-Based Fake News Detection Using NLP and Machine Learning”, *International Journal of Education & Applied Sciences Research*, Volume 13, Issue 1, 2026, pp 54-60. DOI : <https://doi.org/10.5281/zenodo.22268128>

ABSTRACT

The rapid development of digital media and social media networks has increased the dissemination of fake news. Fake news has the ability to deceive readers, shape public opinions, and generate instability in society. It is hard to manually identify fake news due to the vast amount of digital media being posted daily on the internet. This paper proposes a Fake News Detection System that utilizes Machine Learning and Natural Language Processing techniques for the automatic classification of news articles as fake or genuine. The proposed system utilizes text data processing techniques like removal of stop words, tokenization, normalization, etc. The processed data is then converted into numerical data using TF-IDF vectorization. A Logistic Regression classifier is used for the detection of fake news. The proposed system provides the results in the form of a prediction along with a confidence score that indicates the likelihood of the news being genuine or fake. The results indicate that the proposed system has high accuracy in detecting fake news articles. This system has the ability to reduce the dissemination of fake news on the internet.

KEYWORDS: Fake News Detection System, Machine Learning, Natural Language Processing, TF IDF, Logistic Regression.

1. INTRODUCTION

In recent years, online news media and social media have emerged as important channels for disseminating information among people across the world. At the same time, with the growing popularity of these media, fake news and misinformation are spreading rapidly across the world. Fake news is defined as “false or misleading information presented as true news.” If fake news is allowed to spread, it may cause serious problems, such as influencing political decisions, harming people's reputations, and confusing people.

Since thousands of news items are being published daily, it is difficult and time consuming to verify each news item. Machine Learning and Natural Language Processing are important technologies for analyzing a large amount of textual information. Machine Learning and

Natural Language Processing are useful for detecting patterns and characteristics that differentiate between fake news and real news. In this paper, a machine learning-based “Fake News Detection System” is presented.

2. LITERATURE REVIEW

The detection of fake news has become a significant research focus due to the growing impact of social media sites. Various research works have used machine learning algorithms for the classification of news articles. Some studies have used traditional classification algorithms like Naive Bayes, Support Vector Machines, and Decision Trees for analysis. These algorithms have shown promising results in identifying fake news based on writing style.

Recent research has used deep learning algorithms like Recurrent Neural Networks and Long Short-Term Memory networks for the analysis. These algorithms have shown promising results in identifying fake news based on writing style.

Although significant research has been done on the detection of fake news, most systems are facing difficulties in reaching high accuracy rates. Hence, this research proposes a machine learning-based method using TF-IDF and Logistic Regression for efficient classification.

3. OBJECTIVES OF PROPOSED SYSTEM

.The primary objectives of the proposed Fake News Detection System are:

- To create a machine learning model that predicts whether a news article is real or fake.
- To utilize natural language processing techniques for data preprocessing.
- To transform the data into numerical features by using TF-IDF vectorization.
- To utilize a labeled dataset for testing the classification model.
- To present users with a prediction result.
- To help reduce the dissemination of fake news on digital media

4. PROBLEM DEFINITION

The dissemination of fake news through various online media is one of the major issues that society is facing today. People are sharing news without checking its authenticity, which is causing the spread of false information.

Currently, no effective automated system is in place on various online media that can verify the authenticity of news articles. This is causing users to read and spread false information unknowingly.

Manual checking of news is not possible since a huge volume of information is being shared every day. Hence, an automated system is required that can check the authenticity of news articles quickly.

The proposed system, "Fake News Detection System," can be implemented using machine learning algorithms that classify news articles into real or fake news.

5. PROPOSED SYSTEM

The proposed system utilizes a machine learning algorithm along with natural language processing techniques for the automatic detection of fake news.

The proposed system consists of the following steps:

1. Data Collection – A dataset containing both genuine and fake news is collected.
2. Data Preprocessing – Stop words are removed from the data.
3. Feature Extraction – TF-IDF vectorization is used for feature extraction.
4. Model Training – A Logistic Regression model is used for the proposed system.
5. Prediction – The trained model is used for prediction.

6. SYSTEM ARCHITECTURE

The system architecture has three levels:

Presentation Layer

The presentation layer enables users to input news text and display prediction results.

Application Layer

The application layer includes data processing for the text data, feature extraction, and prediction using the machine learning model.

Data Layer

The data layer stores the data used for testing and training the machine learning model.

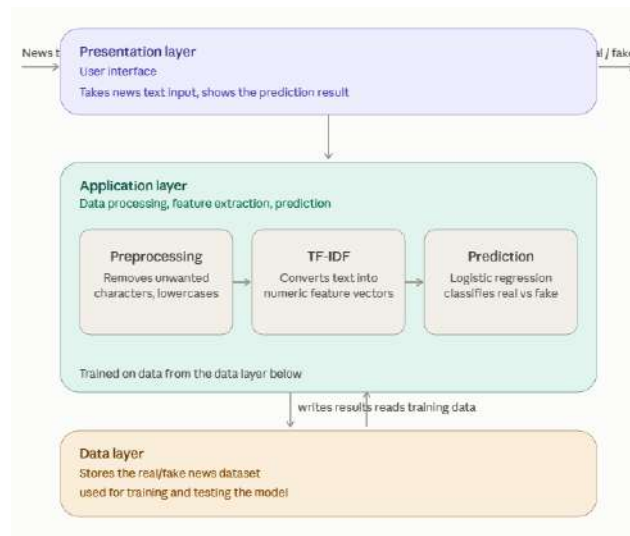


Fig1. System Architecture layers

7. METHODOLOGY

The method used for the Fake News Detection System has different steps.

Firstly, the dataset containing both real and fake news is collected.

Then, the data is pre-processed by removing unwanted characters from the dataset. After that, all the text data is converted into lower case.

Then, the dataset is converted into numerical data using TF-IDF Vectorization. This helps the machine learning model understand the importance of words in the dataset.

Finally, a Logistic Regression model is trained on the dataset. This trained model is used for predicting whether the input news is real or fake.

8. MODULE DESCRIPTION

A. Data Collection Module

This module collects datasets that include real and fake news articles for training the model.

B. Text Processing Module

This module cleans and preprocesses the text data for training the model.

C. Model Training Module

This module trains the machine learning classifier using the cleaned and preprocessed data.

D. Prediction Module

This module analyzes and determines whether the news article is real or fake.

E. Result Display Module

This module displays the result of the prediction and the percentage confidence level. Table1 discusses about various modules and key functions of each module.

Module	Key Functions
Data Collection	Gather datasets
Text Processing	Cleaning and Preprocessing
Model Training	Train ML model
Prediction	Detect whether the news is fake or real
Result Display	Show prediction result

Table 1: Module Functions

9. RESULTS AND DISCUSSION

The system was tested using a dataset with a mix of real and fake news articles. The machine learning model was able to classify most news articles with high accuracy. In our system, users can input any news article, and our machine learning model will be able to determine if it is real or fake.

Table 2: News type and Predictions

News Type	Prediction	Confidence
Political News	Fake	92%
Sports News	Real	88%
Technology News	Real	90%

10. CONCLUSION

The study proposed a machine learning-based fake news detection system that can be used to classify news articles as genuine or fake. The system utilizes natural language processing techniques for text preprocessing and TF-IDF vectorization for feature extraction.

The system utilizes a logistic regression algorithm to analyze the features of the text data to identify patterns related to fake news. The results of the experiment showed that the system is highly accurate and can be used to detect fake news.

The system can be improved in the future by incorporating deep learning techniques to analyze multimedia data such as images and videos.

REFERENCES

- [1] 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.
- [2] 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.
- [3] V. L. Rubin, Y. Chen, and N. J. Conroy, “Deception Detection for News: Three Types of Fake News,” *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.
- [4] N. J. Conroy, V. L. Rubin, and Y. Chen, “Automatic Deception Detection: Methods for Finding Fake News,” *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.
- [5] S. Ruchansky, S. Seo, and Y. Liu, “CSI: A Hybrid Deep Model for Fake News Detection,” *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, pp. 797–806, 2017.
- [6] K. Shu, S. Wang, and H. Liu, “Beyond News Contents: The Role of Social Context for Fake News Detection,” *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 312–320, 2019.
- [7] J. Thorne and A. Vlachos, “Automated Fact Checking: Task Formulations, Methods and Future Directions,” *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3346–3359, 2018.
- [8] S. Vosoughi, D. Roy, and S. Aral, “The Spread of True and False News Online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
- [9] B. Zhou and J. Zafarani, “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities,” *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–40, 2020.
- [10] A. Ahmed, I. Traore, and S. Saad, “Detecting Opinion Spams and Fake News Using Text Classification,” *Security and Privacy*, vol. 6, no. 1, pp. 1–15, 2018.
- [11] Y. Wang, “Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection,” *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 422–426, 2017.

- [12] J. Ma, W. GAO, P. Mitra, S. Kwon, B. J. Jansen, and K. Wong, “Detecting Rumors from Microblogs with Recurrent Neural Networks,” *International Joint Conference on Artificial Intelligence*, pp. 3818–3824, 2016.
- [13] S. Kumar and N. Shah, “FalseInformation on Web and Social Media: A Survey,” *ACM SIGKDD Explorations*, vol. 20, no. 1, pp. 1–19, 2018.
- [14] T. Horne and S. Adali, “This Just In: Fake News Packs a Lot in Title, UsesSimpler, Repetitive Content in Text Body,” *International AAAI Conference on Web and Social Media*, pp. 759–766, 2017.
- [15] D. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some Like It Hoax: Automated Fake News Detection, *arXiv preprint arXiv:1704.07506*, 2017.