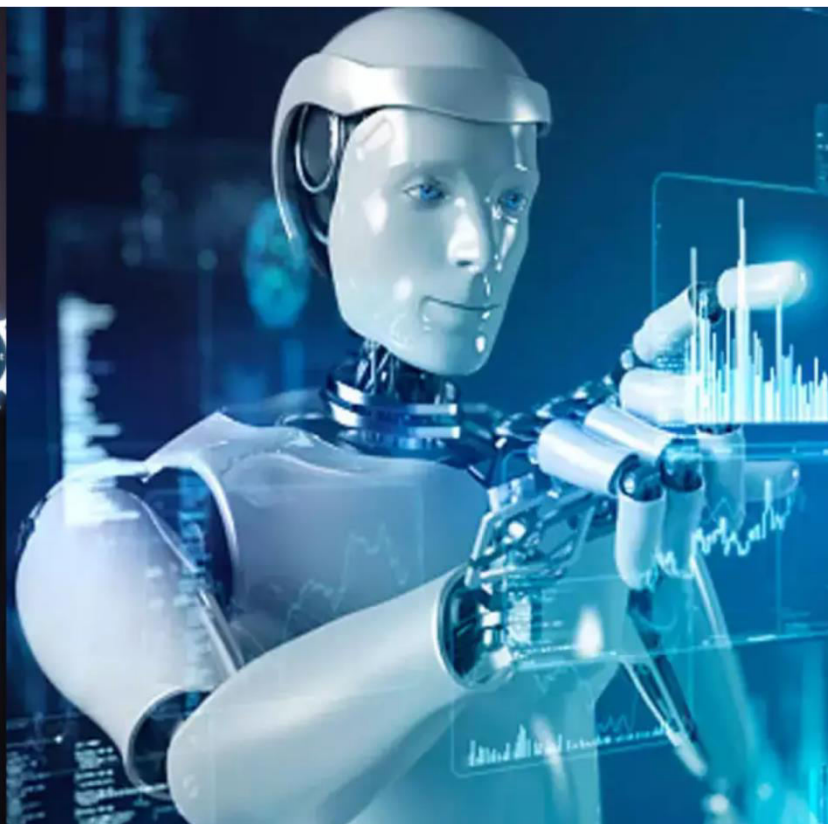




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 10, October 2025

Survey on Fake News Detection using Machine Learning

**Devidas R. Joshi^{*1}, Ganesh B. Morale², Prathmesh P. Gavli³, Sumit S. Shinde⁴,
Prof. Mr. Sarwade S.K⁵**

Department of Information Technology Engineering, Pimpri Chinchwad Education Trust's Pimpri Chinchwad
Polytechnic, Pradhikaran, Nigdi, Pune, India¹⁻⁵

*Corresponding Author

ABSTRACT: With the burgeoning computerized growth, the ecosystem has been plagued with issues like propagation of misinformation and fake news with the social networking sites. This research work is based on the work done toward design of a Fake News detection system via machine learning techniques to automate the process of classification of articles to be genuine or fake . The proposed system uses Natural Language Processing (NLP) techniques for processing of text, and uses supervised learning algorithms which improve the results stochastic . Finally the results show good accuracy , with machine learning as a good method to tackle the issues with detection of fake news and supporting decision making and to counter the problem of misinformation .

KEYWORDS: Fake News, Machine Learning, Text Classification, Natural Language Processing, Python

I. INTRODUCTION

With the spread of internet news sites and social networks, it is easy to find information about different matters, but the propagation of false or deceptive news has also increased. The so-called fake news can modify public opinion, affect the elections and create social problems. Traditional ways of manually verifying truth are laborious and ineffective. Therefore, there is a need for automation of systems that would help timely and effectively detect false news . Machine learning and in particular the application of techniques based on natural language processing provide a solution for efficient analysis of textual content which is concerned with pattern recognition and classification of news .

II. LITERATURE SURVEY / REVIEW

Several works have reported on the detection of fake news using machine learning and deep learning techniques. For example , Zhou et al. (2020) studied text classification by taking advantage of NLP features and achieved a high accuracy by using the Support Vector Machine and Random Forests classifiers. Kaliyar et al. (2019) proposed a hybrid method that combined linguistic features and sentiment analysis, showing that ensemble models enhance the performance of detection. Shu et al. (2017) introduced the dataset FakeNewsNet and stated that, for fake news detection, social context is of importance in addition to text features. Ahmed et al . (2018) worked on feature engineering approaches and concluded that n-grams , TF-IDF and word embeddings are the key features for news detection systems based on machine learning . All these studies show that while traditional ML classifiers offer good base results, the performance of systems can be improved by using a variety of features , content as well as social context .

III. METHODOLOGY / SYSTEM DESIGN

The system devised will work as follows:

1. Data collection : The intent is to collect datasets of news in the following way from recognized news and bogus news repositories (Kaggle, FakeNewsNet).
2. Data preprocessing : Data preprocessing is applied, which includes natural habitat for the text, tokenization , rejection of stop-words and stemming and/or lemmatization .

3. Feature extraction: In this, the numerical representation of the textual data is obtained by means of TF - IDF vectors and word embedding.
4. Model selection : Supervised learning is done by all such algorithms as Logistic Regression, Naïve Bayes , Random Forest and Support Vector Machines (SVM).
5. Training and Testing : The datasets worked upon are segregated into two parts , the training (80%) and the testing (20%) . The models are trained on labelled training data collected by them and subsequently tested on accuracy , precision, recall and F1 metrics .
6. Prediction: The new articles are classified as real and bogus on the basis of the trained model.

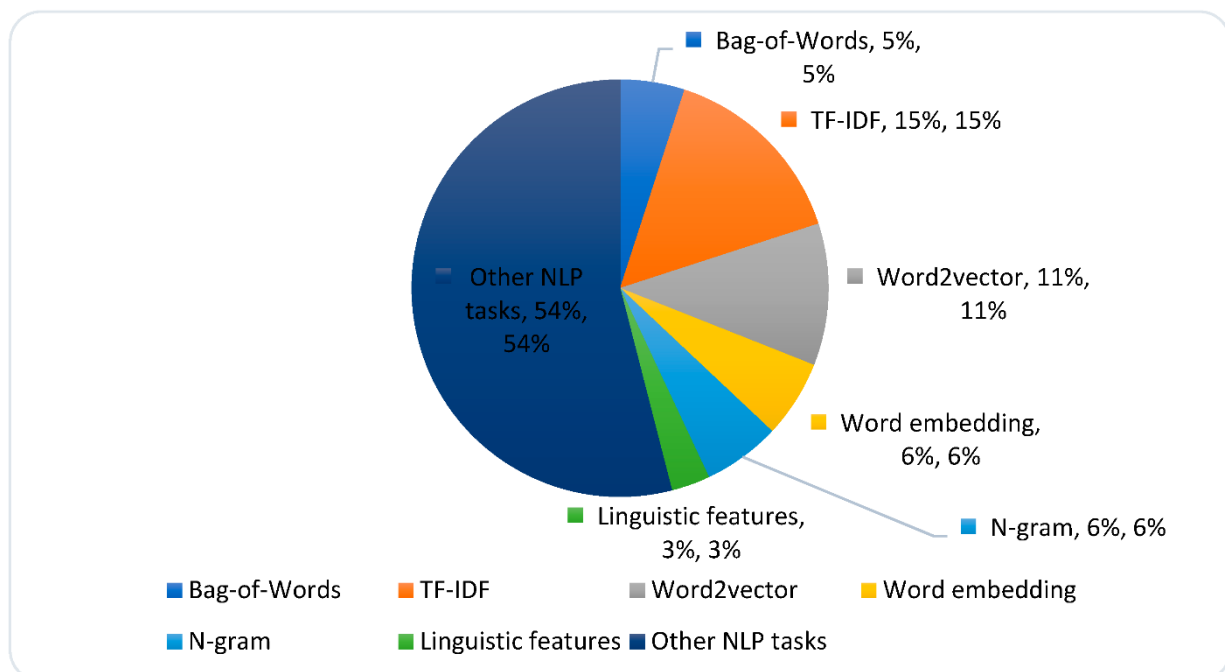
IV. IMPLEMENTATION / WORKING

1. Environmental Setup : The languages used were Python and Jupyter Notebook using the Pandas , NumPy , Scikit-learn, NLTK and Matplotlib libraries.
2. Data Preparation: The data was processed to get rid of special symbols, HTML tags and unnecessary symbols.
3. Feature Engineering: TF-IDF was used for the vectorisation of the text information, converted to labelled features.
4. Model Training: Logistic Regression and Random Forest were used as classifier. Test and Evaluation: The models were tested using unknown data.

V. RESULT AND DISCUSSION

The experimental results show that the Random Forest classifier gives better values than with the other models in terms of both accuracy and in F1 -score. Then the confusion matrix offers a low false-positive rate, which is very important for practical applications. The system will be able to indicate the fake news in the system very successfully, and it is very important to provide the users with reliable information. However, the performance may vary according to the quality of the dataset, and the meaningfulness of the features in connection with this particular model.

VI. SYSTEM ARCHITECTURE



The system architecture of the Fake News Detection project is designed to identify and classify online news articles as real or fake using various Natural Language Processing (NLP) techniques and machine learning models. The process begins with collecting news data from trusted and untrusted sources, followed by preprocessing steps such as

tokenization, stop-word removal, and lemmatization to clean and standardize the text. In the next stage, important textual features are extracted using techniques like TF-IDF, Word2Vec, Word Embeddings, N-grams, and Bag-of-Words, as shown in the figure. These methods help the model understand both the meaning and structure of the news content. The extracted features are then fed into machine learning or deep learning algorithms, which are trained to learn patterns that differentiate fake news from real news. Finally, the trained model classifies new or unseen articles and provides the output, helping detect misleading or false information effectively.

VII. CONCLUSION AND FUTURE SCOPE

1. This papershows that machine learning is effective in detecting fake news. An effective method of checking the news has been put in place . Future proposals:
2. Implement Artificial Intelligence LSTM and deception detection BERT has been tested in deep learning models to obtain abetter understanding of context .
3. Use social context and social networking has been used to provide better detection of false news.
4. A real-time production of fake news detection has been developed for social networking.

REFERENCES

1. Shu K., Sliva A., Wang S., Tang J. and Liu H. (2017). FakeNewsNet : A Data Repository with News Content , Social Context and Spatio-Temporally Rich Information for Fake News Research. arXiv Prepr. arXiv :1709.01286
2. Zhou, X., and Zafarani, R. (2020). A Survey on Fake News: Fundamental Theories, Detection Methods and Opportunities. ACM Comput. Surv. (CSUR) 53, 1–40.
3. Kaliyar , R. C., Goswami, A., and Narang, P. (2019). Fake News Detection using Machine Learning. Int. J. Inf. Technol.
4. Ahmed, H., Traore, I., and Saad, S., (2018). Online Fake News Detection using N-gram Analysis and Machine Learning Techniques. Soc. Netw. Anal. Min. 8, 1–15.
5. Kaggle dataset we can handle Fake News, <https://www.kaggle.com/datasets> .



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details