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 5, Issue 12, April 2025
Impact Factor: 7.67

Identifying Fake News using Machine Learning Models

Anusha Mohan Amrutkar[1] and Prof. N. V. Alone[2]
Department of Computer Engineering[1,2]
Gokhale Education Society's,
R. H. Sapat College of Engineering Management Studies and Research, Nashik

Abstract: *In today's digital era, where the internet connects the world, people increasingly rely on various online platforms for accessing news. With the widespread use of social media platforms such as Facebook, Twitter, and others, news—whether real or fake—can spread to millions of users within moments. The rapid dissemination of fake news can have serious consequences, including the creation of biased opinions and the spread of misinformation.*

This project focuses on the detection of fake news using advanced technologies. The dataset for this study was provided by an external organization. The objective is to perform binary classification of various online news articles using concepts from Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML).

The system uses a Decision Tree Classifier to classify news articles as either fake or real. To prepare the text data for machine learning models, different feature engineering methods have been employed, including the Bag of Words (BoW) model and Word Embedding techniques. These methods convert textual data into numerical feature vectors, which are then fed into machine learning algorithms for classification.

Various combinations of features and classification algorithms are explored to determine the most effective approach for fake news detection. The model that achieves the best performance—in terms of feature extraction method and classification algorithm—is selected for final prediction, ensuring accurate identification of fake or real news.

Keywords: News Identification dataset, Deep Learning, Machine Learning, Classification.

I. INTRODUCTION

In today's digital era, where the internet connects the world, people increasingly rely on various online platforms for accessing news. With the widespread use of social media platforms such as Facebook, Twitter, and others, news—whether real or fake—can spread to millions of users within moments. The rapid dissemination of fake news can have serious consequences, including the creation of biased opinions and the spread of misinformation. This project focuses on the detection of fake news using advanced technologies. The dataset for this study was provided by an external organization. The objective is to perform binary classification of various online news articles using concepts from Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML).The system uses a Decision Tree Classifier to classify news articles as either fake or real. To prepare the text data for machine learning models, different feature engineering methods have been employed, including the Bag of Words (BoW) model and Word Embedding techniques. These methods convert textual data into numerical feature vectors, which are then fed into machine learning algorithms for classification. Various combinations of features and classification algorithms are explored to determine the most effective approach for fake news detection. The model that achieves the best performance—in terms of feature extraction method and classification algorithm—is selected for final prediction, ensuring accurate identification of fake or real news.

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 5, Issue 12, April 2025
Impact Factor: 7.67

II. PURPOSE

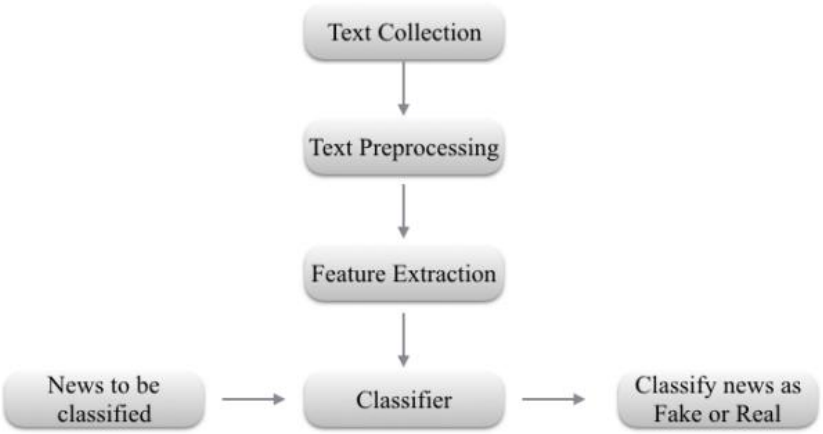
This project focuses on the development of a fake news classification system utilizing various feature extraction methods and classification algorithms. These algorithms include Support Vector Machine (SVM), Logistic Regression, Gradient Boosting, XGBoost, Decision Tree, and Random Forest. The goal is to identify the best-performing algorithm, which will then be used to classify news as either fake or real.To build a real-time application, the algorithm must be continuously updated with the most recent data. Given that the data varies in size, proper data cleaning is crucial to improve the model's accuracy. To optimize the results, we apply a combination of feature extraction techniques, such as the Bag of Words (BoW) model and Word Embedding. By experimenting with different algorithms and feature extraction methods, we aim to identify the approach that yields the highest classification accuracy for fake news detection.

III. OBJECTIVE OF SYSTEM

- Data Collection and Analysis
- Data Preprocessing
- Text Feature Extraction
- Applying Different Classification Algorithms
- Selecting the Best Classification Algorithm and Feature Extraction Method

IV. SYSTEM ARCHITECTURE

In future works, we intend to use highly sophisticated classifying approach, like deep learning with sentiment analysis also and consider many text features like publisher, urls etc., which may increase the accuracy of the classification of news as fake or real. Automatic fake news detection may be done using the latest news and training the model regularly to get the best results. So this can be used as a filter to upload the news.



User Registration:
Here User has to register with required parameters such as name, mobile, password and hardware ID.
User Login:
After user registration done successfully user can login to the system
Predict News
Feed Back Generation

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 5, Issue 12, April 2025
Impact Factor: 7.67

Algorithm & Mathematical Model
S={I, O, P, S, C, P, Ad, Q, G,H/w,S/w, Failure, Success} Where
S=System
C= Check News U=User
Ad=Admin
G= Detection.
Procedures {P}= {Pr, Cc, Qid, Amt} Where,
Pr= Check News
Qid= Find News
O is Output of system

VI. CONCLUSION

In this paper three different feature extraction methods like Count Vectorizer, TFIDF Vectorizer, Word Embedding has been used. And also different classification algorithms like Linear SVC, Logistic Regression Classifier, Decision Tree Classier, Random Forest Classifier, XG-BOOST Classifier, Gradient Boosting Classifier have been used to classify the news as fake or real. By using the classification algorithms we got highest accuracy with SVM Linear classification algorithm and with TF-IDF feature extraction with 0.94 accuracy. Even though we got the same accuracy with Neural Network with Count Vectorizer, Neural Networks and take more time to train and its complex so we used Linear SVC which is not so complex and takes less time to computer .

REFERENCES

[1] U. A. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi, and A. Ahmadian, “Cloud-based email phishing attack using machine and deep learning algorithm,” Complex Intell. Syst., pp. 1–28, Jun. 2022.
[2] D. C. Le and A. N. Zincir-Heywood, “Machine learning based insider threat modelling and detection,” in Proc. IFIP/IEEE Symp. Integr. Netw. Service Manag. (IM), Apr. 2019, pp. 1–6.
[3] P. Oberoi, “Survey of various security attacks in clouds based environments,” Int. J. Adv. Res. Comput. Sci., vol. 8, no. 9, pp. 405–410, Sep. 2017.
[4] A. Ajmal, S. Ibrar, and R. Amin, “Cloud computing platform: Performance analysis of prominent cryptographic algorithms,” Concurrency Comput., Pract. Exper., vol. 34, no. 15, p. e6938, Jul. 2022.
[5] U. A. Butt, R. Amin, M. Mehmood, H. Aldabbas, M. T. Alharbi, and N. Albaqami, “Cloud security threats and solutions: A survey,” Wireless Pers. Commun., vol. 128, no. 1, pp. 387–413, Jan. 2023.
[6] Mykhailo Granik and Volodymyr Mesyura, "Fake news detection using naive bayes classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900-903, 2017.
[7] Cédric Maigrot, Ewa Kijak and Vincent Claveau, "Fusion par apprentissage pour la détection de fausses informations dans les réseaux sociaux", Document numerique, vol. 21, no. 3, pp. 55-80, 2018.
[8] Cristina M Pulido, Laura Ruiz-Eugenio, Gisela Redondo-Sama and Beatriz Villarejo-Carballido, "A new application of social impact in social media for overcoming fake news in health", International journal of environmental research and public health, vol. 17, no. 7, pp. 2430, 2020.
[9] Juan Ramos et al., "Using tf-idf to determine word relevance in document queries", Proceedings of the first instructional conference on machine learning, vol. 242, pp. 133-142, 2003.
[10] Bernhard Scholkopf and Alexander J Smola, "Learning with kernels: support vector machines regularization optimization and beyond", Adaptive Computation and Machine Learning series, 2018.