e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

FAKE NEWS DETECTION WITH MACHINE LEARNING ALGORITHMS

Himani Rane*1

*1UG Student, Department Of Information Technology, B. K. Birla College
(Empowered Autonomous Status), Kalyan, India.

ABSTRACT

Fake news has become a pervasive issue in the digital age and the need for effective methods to detect and combat it has grown significantly. This research focuses on the development and evaluation of a fake news detection system using TfidfVecorizer, svm, Logistic Regression and Random Forest Classifier model. The primary goal is to assess the accuracy of the detection system.

Keywords: Fake News, True News, Machine Learning, Vectorization.

I. INTRODUCTION

In recent years, the proliferation of the fake news has emerged as a formidable challenge, threatening the integrity of information dissemination, the functioning of democratic societies and the credibility of journalism. Fake news, characterized by the dissemination of false or misleading information presented as legitimate news, can have far-reaching consequences, from misleading the public and exacerbating social tensions to undermining trust in media and institutions.

The proliferation of fake news in the digital age poses a significant challenge to the integrity of information and its impact on society. The spread of deceptive or false information through various online platforms has the potential to influence public opinion, elections, and even public health. To combat this growing issue, the development of effective fake news detection models is crucial. This research paper introduces a comprehensive approach to fake news detection utilizing a combination of techniques, including TF-IDF vectorization, train-test splitting, logistic regression, and accuracy score evaluation.

II. METHODOLOGY

Data Collection: Obtain two separate datasets: one containing true or real news articles and the other containing fake news articles. Ensure that both datasets are well-annotated to distinguish between real and fake news.

Data Preprocessing:

Text Cleaning: Apply text cleaning techniques to remove unnecessary characters, symbols, HTML tags, and special characters from the text in both datasets.

Tokenization: Tokenize the text in both datasets, splitting them into individual words or tokens.

Stop-Word Removal: Eliminate common stop words from both datasets to prepare them for further processing.

Lowercasing: Convert all text to lowercase for consistency.

Stemming or Lemmatization: Consider applying stemming or lemmatization to both datasets to reduce words to their root form.

Feature Extraction with TF-IDF Vectorization: Apply TF-IDF vectorization separately to both the true news dataset and the fake news dataset. This process will create TF-IDF vectors for each article in both datasets, quantifying the importance of words within the context of each article.

Combining Datasets and Train-Test Split: After TF-IDF vectorization, combine the two datasets, creating a unified dataset with labeled entries (real or fake news). Split the unified dataset into a training set and a testing set. The training set is used to train the machine learning models, and the testing set is used to evaluate the models' performance on unseen data.

Model Selection and Training: Initialize and train three classification models: Support Vector Machine (SVM), Logistic Regression, and Random Forest Classifier. Train each model using the training data to learn the relationships between the TF-IDF vectors and the corresponding labels (real or fake news).

Model Evaluation: Evaluate the performance of each model using the following metrics:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Accuracy: To measure the overall correctness of predictions.

Precision: To measure the accuracy of positive predictions (fake news).

Recall: To measure the coverage of actual positive cases (fake news).

F1-score: A combination of precision and recall to balance model performance.

Compare how well each model classifies both real and fake news articles in the testing dataset.

III. EXPERIMENTAL RESEARCH

As mentioned before after setting up the experiments and training the models on the dataset the following are the result.

Table 1: Classification Report

Model	Accuracy	Recall	Precision	F1_score
Logistic Regression	0.99	0.99	0.99	0.99
Svm	0.98	0.98	0.99	0.98
Random Forest Classifier	0.995	0.99	1.0	0.99

Table 1 shows the classification report of all the models, the table displays the Accuracy, Precision, and Recall of all the models, calculated using the test data that was split at the beginning of the experiment.

IV. CONCLUSION

In conclusion, our research contributes to the ongoing efforts to combat the spread of fake news by providing a systematic evaluation of machine learning models. By implementing TF-IDF vectorization and considering various classification algorithms, we demonstrate the potential for creating reliable and effective fake news detection tools. As technology and misinformation continue to evolve, research in this domain will remain critical in ensuring the integrity of information in the digital age.

V. REFERENCES

[1] N. F. Baarir and A. Djeffal, “Fake News detection Using Machine Learning,” in 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), Boumerdes, Algeria: IEEE, Feb. 2021, pp. 125–130. doi: 10.1109/IHSH51661.2021.9378748.

[2] J. Shaikh and R. Patil, “Fake News Detection using Machine Learning,” in 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), Gunupur Odisha, India: IEEE, Dec. 2020, pp. 1–5. doi: 10.1109/iSSSC50941.2020.9358890.