

## DETECTING ONLINE FAKE NEWS USING SUPERVISED MACHINE LEARNING ALGORITHMS

Chijioke Emmanuel Okoro and Ifeoma Chinyere Eze

Department of Computer Science, Rivers State University, Port Harcourt, Nigeria

Department of Information Technology, Rivers State University, Port Harcourt, Nigeria

### Abstract

Fake news, defined as intentionally false information circulated through digital platforms, poses a significant threat to information integrity, particularly on social media. Given its persuasive nature, it is critical to develop efficient methods for detecting fake news and promoting more responsible consumption of online content. This paper proposes a model that leverages supervised machine learning algorithms to detect fake news in online sources. The dataset used in this study, "fake\_or\_real\_news," was preprocessed through feature extraction, narrowing down the data to two columns: text and labels. The label column was further processed to mark true news as "REAL." The study applied three different supervised machine learning algorithms for model training: Logistic Regression, Support Vector Classifier (SVC), and Multinomial Naive Bayes (MultinomialNB). The models were evaluated based on accuracy, with Logistic Regression achieving the highest accuracy at 91.9%, followed by MultinomialNB at 88.5%, and SVC at 86.8%. Based on the results, the study recommends using Logistic Regression as the most effective algorithm for detecting fake news in online environments. This approach offers valuable insight into how machine learning can be employed to combat misinformation.

**Keywords:** Fake news detection, supervised learning, machine learning, Logistic Regression, social media.

### 1. Introduction

Fake news in Nigeria has caused a lot of damages especially in political era, spreading false information across the internet in order to make critics, cause religious and ethnic division, damage reputations as well as drawing more persons to the website. Some persons do spread fake information on their websites in order to get more persons visiting their website. Fake news refers to intentionally and verifiably false stories that are largely disseminated through social media networks. It can be very persuasive and therefore makes it necessary to develop strategies to identify and critically assess news read and circulated on social media. False news reached more people than the truth; the top 1% of false news cascades diffuse to between 1000 and 100,000 people, whereas the truth rarely diffuses to more than 1000 people. Falsehood also diffuses faster than the truth. The relevance of fake news has increased in post-truth politics. For media outlets, the ability to attract viewers to their websites is necessary to generate online advertising revenue; publishing a story with false content that attracts user's benefits advertisers and improves ratings. Fake news has recently raised a lot of concerns because of its prevalent impact. The problem

is not only unique to online environments; it is also present in the conventional media. The internet has enabled a whole new way to publish, share and consume information and news with very little regulation or editorial standards. Many people now get news from social media sites and networks and often it can be difficult to tell if stories are credible or not. Information overload and a general lack of understanding about how the internet works, has also contributed to an increase in fake news or hoax stories. Social media sites can play a big part in increasing the reach of these types of stories.

The proliferation of Fake News on social media has been a source of widespread concern. One of the main approaches to automatically detect this type of news is based on crowd signals, i.e., opinions manifested by social media users concerning whether the news are fake or not. Although promising, this approach depends on information that is not always available: the explicit opinion of the users about the news to be checked [1]. During natural disasters or crises, users on social media tend to easily believe contents of postings related to the events, and retweet the postings with hope them of reaching to many other users. Unfortunately, there are malicious users who understand the tendency and post misinformation such as spam and fake messages with expectation of wider propagation [2]. This paper provides a model for detecting online fake news using supervised machine learning algorithms.

## **2. Related Work**

The paper “Fake and spam messages detecting misinformation during natural disasters on social media” conducted a case study of 2013 Moore Tornado and Hurricane Sandy [2]. Concretely, they understand behaviors of these malicious users, analyze properties of spam, fake and legitimate messages, proposed flat and hierarchical classification approaches, and detect both fake and spam messages with the ability to distinguish between them. Their experimental results show that their proposed approaches identify spam and fake messages with 96.43% accuracy and 0.961 F-measure.

The paper “The spread of true and false news online” investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017 [3]. The data they used comprises of 126,000 stories tweeted by 3 million people more than 4.5 million times. They classified news as true or false using information from six independent fact-checking organizations that exhibited 95 to 98% agreement on the classifications. They also found that false news was more novel than true news, which they suggest that people were more likely to share novel information. It was also reveal in the study that robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it.

The paper “Automatically Identifying Fake News in Popular Twitter Threads” presented by [4] developed a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two

credibility-focused Twitter datasets: CREDDBANK, a crowdsourced dataset of accuracy assessments for events in Twitter and PHEME, a dataset of potential rumors in Twitter and journalistic assessments of their accuracies. They applied this method to Twitter content sourced from Buzz Feed's fake news dataset and show models trained against crowdsourced workers outperform models based on journalists' assessment and models trained on a pooled dataset of both crowdsourced workers and journalists. All three datasets, was aligned into a uniform format. They use a feature analysis in identifying features that are most predictive for crowdsourced and journalistic accuracy assessments, results of which are consistent with prior work. The paper, concluded with a discussion contrasting accuracy and credibility and why models of non-experts outperform models of journalists for fake news detection in Twitter.

The paper "Fake news detection on social media: A data mining perspective" presented a comprehensive review of detecting fake news on social media, including fake news characterization on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets [5].

The paper "Fake news detection on social media via implicit crowd signals" proposed a crowd signal-based method that does not demand the users' explicit opinion to detect fake news [1].

The proposed method infers the users' opinion from their news spreading (publication/propagation) behaviour. Preliminary experiments carried out with two real-world datasets provided evidence that the proposed method can detect fake news without demanding the explicit opinion of users about the news and without compromising the classification results obtained by the state-of-the-art crowd signal-based method.

### 3. Design methodology

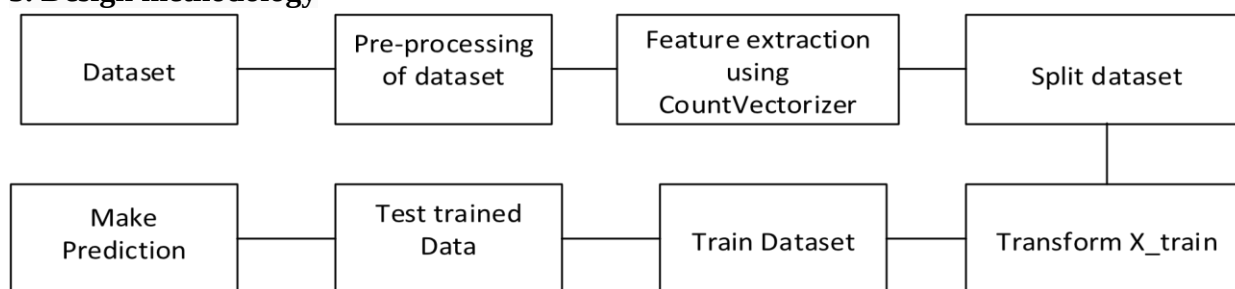


Figure 1: Architecture of the proposed system

This system uses a dataset called fake\_or\_real\_news. The dataset was downloaded from kaggle.com with 4 columns which are unnamed, title, text and label. The dataset was preprocessed to be text and label. The text column contains both real and fake news information while the label column shows if the news is real or fake. The label column was also preprocessed to be either 1 or 0, where 1 represents that news is real and 0 represents

# Artificial Intelligence, Machine Learning, and Data Science Journal

## Research Article

that the news is fake. The new dataset (text, label) was split into x variable and y variable, where x holds the text column and y holds the label column. The x variable was parsed to remove words, called tokenization, this words need to be encoded as integers or floating point values for use as input in training a machine learning algorithm. This whole process is called feature extraction which we used countVectorizer in achieving it. The dataset was trained using 3 machine learning algorithms which are MultinomialNB from Naïve\_Bayes, Support Vector Classifier and Logistic Regression. We used these three models in making predictions and scoring their accuracy.

| Unnamed: 0 |       | title                                             | text                                              | label |
|------------|-------|---------------------------------------------------|---------------------------------------------------|-------|
| 0          | 8476  | You Can Smell Hillary's Fear                      | Daniel Greenfield, a Shillman Journalism Fello... | FAKE  |
| 1          | 10294 | Watch The Exact Moment Paul Ryan Committed Pol... | Google Pinterest Digg LinkedIn Reddit Stumbleu... | FAKE  |
| 2          | 3608  | Kerry to go to Paris in gesture of sympathy       | U.S. Secretary of State John F. Kerry said Mon... | REAL  |
| 3          | 10142 | Bernie supporters on Twitter erupt in anger ag... | — Kaydee King (@KaydeeKing) November 9, 2016 T... | FAKE  |
| 4          | 875   | The Battle of New York: Why This Primary Matters  | It's primary day in New York and front-runners... | REAL  |

Figure 2: showing original dataset downloaded from kaggle.com created by Raluca Chitic

### 4. Result and discussion

In this paper, a machine learning model was trained to detect if a news is real or fake. This model uses fake\_or\_real\_news dataset which have 4 columns but was preprocessed to a dataset with two columns (text, label). This new dataset was divided into x and y variable where x variable holds both fake and real news. The x variable was fit transformed using CountVectorizer class which was gotten from feature extraction. The variable x and y were further divided into x\_train, x\_test, y\_train and y\_test. This x\_train and y\_train where being fitted or trained using four machine algorithms which are LogisticRegression, Support Vector Classifier and MultinomialNB from Naïve\_Bayes. These 3 algorithms were tested and scored based on accuracy, LogisticRegression had an accuracy of about 91.9%, Support Vector Classifier had an accuracy of about 86.8% and MultinomialNB had an accuracy of about 88.5%. The model highest accuracy is the Logistic Regression which was saved and used in detection of fake news and real news.

# Artificial Intelligence, Machine Learning, and Data Science Journal

## Research Article

|   | text                                              | REAL |
|---|---------------------------------------------------|------|
| 0 | Daniel Greenfield, a Shillman Journalism Fello... | 0    |
| 1 | Google Pinterest Digg Linkedin Reddit Stumbleu... | 0    |
| 2 | U.S. Secretary of State John F. Kerry said Mon... | 1    |
| 3 | — Kaydee King (@KaydeeKing) November 9, 2016 T... | 0    |
| 4 | It's primary day in New York and front-runners... | 1    |

Figure 3: showing Training data which was been used in feeding the 3 machine learning algorithms

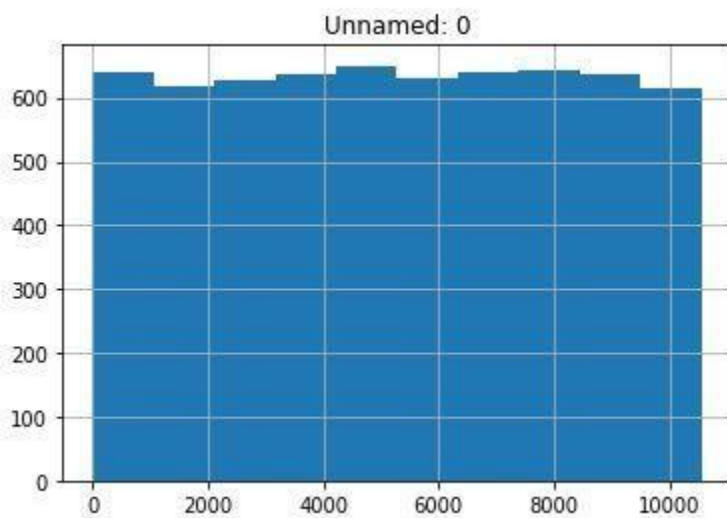


Figure 4: showing the histogram of the entire dataset

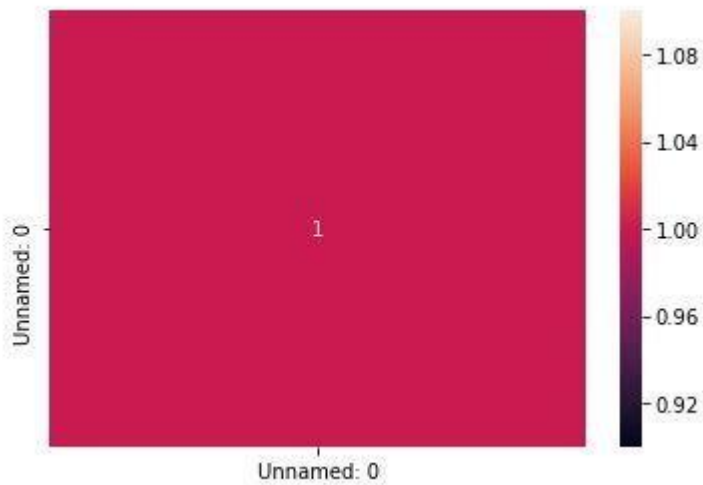


Figure 5: showing correlation matrix of the dataset

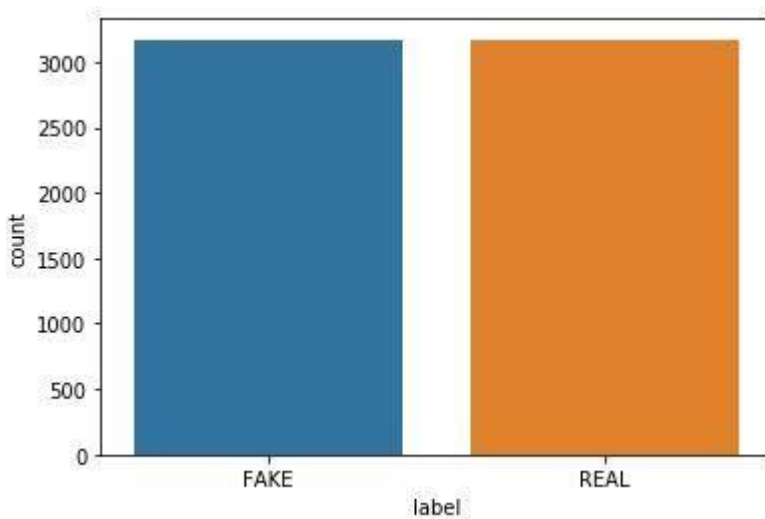


Figure 6: Showing a Count plot of real and fake news

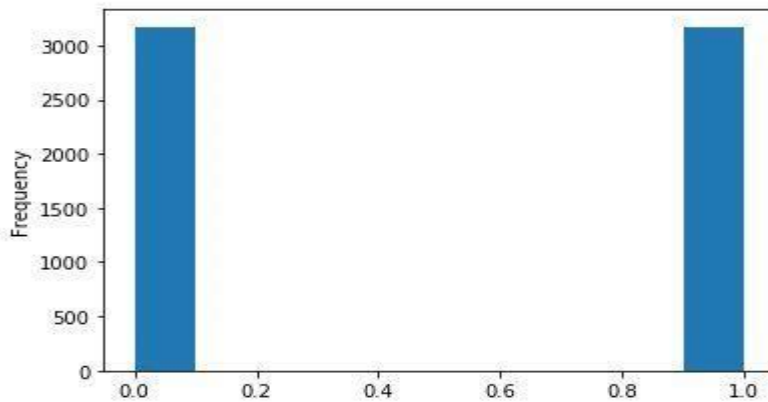


Figure 7: showing a histogram of real and fake news

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.82      | 0.94   | 0.87     | 1066    |
| 1            | 0.92      | 0.78   | 0.85     | 1025    |
| accuracy     |           |        | 0.86     | 2091    |
| macro avg    | 0.87      | 0.86   | 0.86     | 2091    |
| weighted avg | 0.87      | 0.86   | 0.86     | 2091    |

Figure 8: showing a classification report of support vector classifier

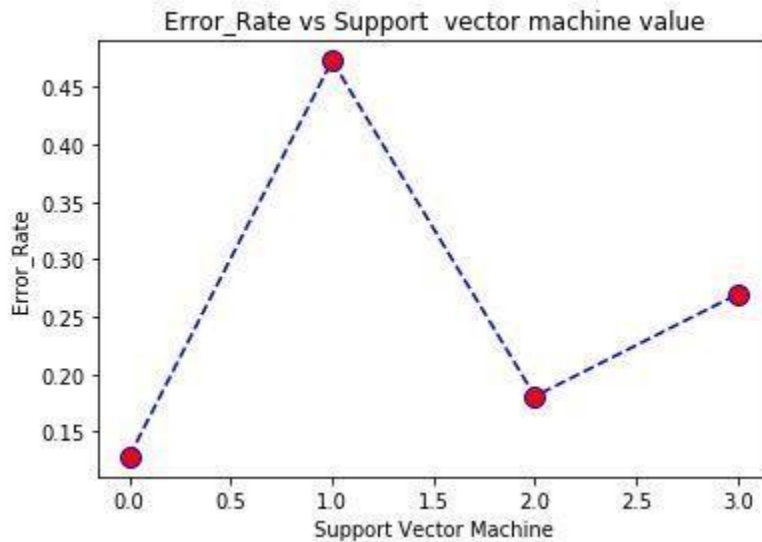


Figure 9: showing error rate vs support vector classifier

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.92   | 0.92     | 1066    |
| 1            | 0.92      | 0.92   | 0.92     | 1025    |
| accuracy     |           |        | 0.92     | 2091    |
| macro avg    | 0.92      | 0.92   | 0.92     | 2091    |
| weighted avg | 0.92      | 0.92   | 0.92     | 2091    |

Figure 10 showing classification report on Logistic Regression algorithm

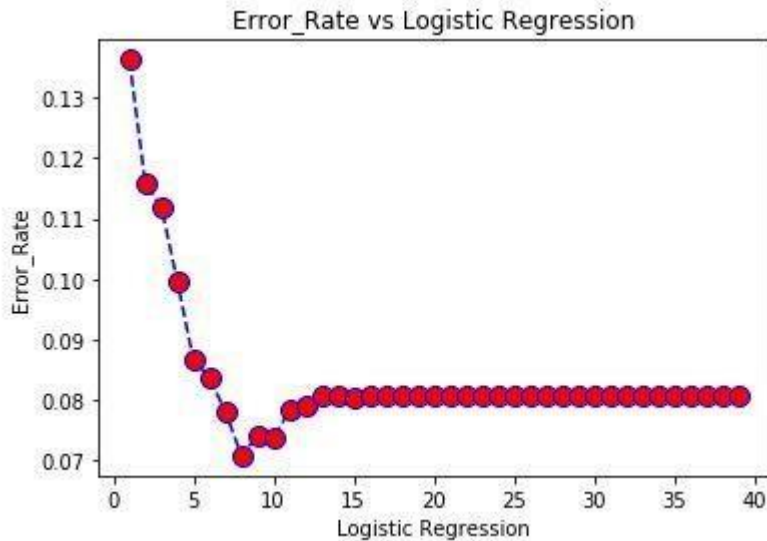


Figure 11 showing error rate vs Logistic Regression

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.85   | 0.88     | 1066    |
| 1            | 0.85      | 0.93   | 0.89     | 1025    |
| accuracy     |           |        | 0.89     | 2091    |
| macro avg    | 0.89      | 0.89   | 0.89     | 2091    |
| weighted avg | 0.89      | 0.89   | 0.89     | 2091    |

Figure 12 showing classification report of MultinomialNB classifier

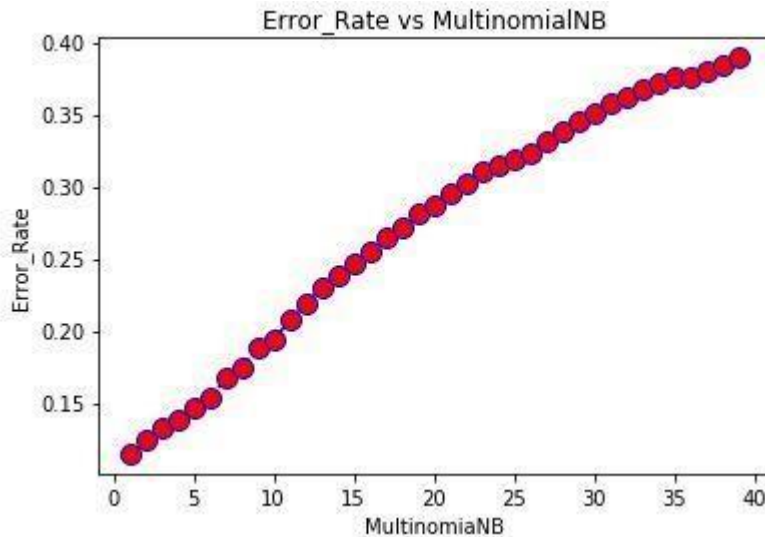


Figure 13 showing error rate vs MultinomialNB classifier

### 5. CONCLUSION AND FUTURE SCOPE

In this paper three machine learning algorithms were used to train and analyse the model. A dataset which contains both real and fake news information was used in feeding and training the model. After testing for accuracy, LogisticRegression had the highest accuracy which is approximately 91.9%, while MultinomialNB had approximately 88.5% accuracy and Support Vector Classifier had 86.8% accuracy. Logistic Regression was saved and tested in detecting fake news; thus presenting the model with a better performance in detecting fake news. This paper can further be extended by using convolution neural network in training to take both pictures and text in feeding the network. It can also be extended into a real system by using it as a chrome browser extension.

### REFERENCES

P. Mario, S. Freire, “Fake news detection on social media via implicit crowd signals”

Proceedings of the 25th Brazilian Symposium on Multimedia and the Web, pp. 521-524. 2019.

M. Rajdev and K. Le, “Fake and Spam Messages: Detecting Misinformation During Natural Disasters on Social Media,”. IEEE, pp. 17–202015.

# Artificial Intelligence, Machine Learning, and Data Science Journal

## Research Article

- S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
- C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," pp. 208–215, *IEEE*, Nov. 2017.
- K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *SIGKDD Explorations*, vol. 19, no. 1, pp. 22–36, 2017.