

# An Improved Classification Model for Fake News Detection in Social Media

**Bodunde Akinyemi**

Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria  
E-mail: bakinyemi@oauife.edu.ng

**Oluwakemi Adewusi and Adedoyin Oyebade**

Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria  
E-mail: odunkemi@gmail.com, sacnet2010@yahoo.com

Received: 30 November 2019; Accepted: 25 December 2019; Published: 08 February 2020

**Abstract**—Fake news dissemination is a critical issue in today's fast-changing network environment. Existing classification models for fake news detection have not completely stopped the spread because of their inability to accurately classify news, thus leading to a high false alarm rate. This study proposed a model that can accurately identify and classify deceptive news articles content infused on social media by malicious users. The news content, social-context features and the respective classification of reported news was extracted from the PHEME dataset using entropy-based feature selection. The selected features were normalized using Min-Max Normalization techniques. A predictive fake news detection model was formulated as a stacked ensemble of three algorithms. The model was simulated and its performance was evaluated by benchmarking with an existing model using detection accuracy, sensitivity, and precision as metrics. The result of the evaluation showed a higher 17.25% detection accuracy, 15.78% sensitivity, but lesser 0.2% precision than the existing model. Thus, the proposed model detects more fake news instances accurately based on news content and social content perspectives. This indicates that the proposed classification model has a better detection rate, reduces the false alarm rate of news instances and thus detects fake news more accurately.

**Index Terms**—Fake news, classification, stacking ensemble, news instances, news content, social-context features, social media.

## I. INTRODUCTION

Over the years, there has been an increase in social media-generated fake news. The social media services have facilitated sharing, production and searching of information by people at an unprecedented level. For example, in US over 65% of the Adult read news from Facebook, Twitter and other social media platform [1]. Also, in 2018, Nigeria had 92.3 million internet users which is projected to grow to 187.8 million internet users in 2023 [2], out of which 72% of Nigerians with access to

internet were visiting social media sites [3]. The social media as a medium used by people for news updates and sharing is a double-edged sword. Despite all the advantages provided by the social media in sharing of information, the quality or veracity of news is much lower than the traditional news source. It has become extremely difficult for individuals reading the news on social media to decipher fake or real news. Social networks offer platforms in which the contents can be relayed by any users without significant third-party filtering, fact-checking and editorial judgment [4] [5].

It has been noted that fake news spread more than the real news on social media. Fake news also called distorted fact, rumor, mis-information or bias news are news articles that are intentionally and verifiably false being spread via traditional news media or social media [6][7]. Fake news misleads readers, have a serious harmful effect, and can break the authenticity balance of the news ecosystem. It intentionally persuades consumers to accept biased or false beliefs' to communicate a manipulator's agenda and gain influence and fake news changes the way people interpret and respond to real news [8].

Fake news is a global threat to world peace, economic growth and security. Research has shown that 46% of the fake news from the social media site is about US politics and election others are for anti-social behaviour such as cyberbullying, propaganda, hate crimes and for radicalization and recruitment of individuals into terrorism organization such as ISIS [9] [10] [11]. For example, the news report about Steve Job's heart attack which had a resultant effect on the stock of Apple Inc. [5], the engulfment of the White House [12], endorsement of Donald Trump by Pope Francis for the United States President [13] etc. Also, in 2015, there are various reported cases of fake news in Nigeria media. Some of the headlines are, "*Nnamdi Kanu Apologies to President Buhari*", "*President Buhari closes down the Aso Rock Villa Church*", "*Godsday Orubebe defects from PDP to APC*", "*Owa Obokun of Ijesaland has Passed Away*" etc. All of these news headlines and contents are intentionally created to misinform the people and cause political

tension.

Fake news are a constant menace to the social media users and the populace at large. Its resultant adverse psychological effects have caused anxiety, hatred and violence among the citizenry. The upsurge of fake news on social media has significantly deteriorated the Quality of Services (QoS) of the new media. This has attracted the urgent need for verifying the veracity of news content to decipher fake news from news content. To this effect, several solutions have been suggested to address the dissemination of fake news on the social media. Despite all these approaches, the spread of fake news have not completely stopped. Existing solutions lack the effective capability to accurately classify news. This degeneration has given rise to the research on the detection of fake news efficiently and accurately, hence this study.

The prediction of the chances of a particular news article to be intentionally deceiving is a challenging task. The traditional approaches which mostly rely on using lexical features or manual annotation by third parties to notify users, have only very limited success, due to the rate at which fake news is being generated and the rate at which it goes viral in few minutes, thus the need for automatic, computational fake news detection system [14] [15].

Several automated models have been developed for the detection of fake news on social media which either focuses on the News context models or Social context models. News content models focus on the content of the news in the detection of fake news while the Social context models focus on the social features and signals which is usually the interaction or response of news reader to the news content. Such automated model includes fake news characterizations on psychology and social theories, algorithms from a data mining approach. From the Data mining perspective, a general classification model for fake news detection includes feature extraction and model construction phases. The feature extraction phase aims to represent news content and related auxiliary information in a formal mathematical structure, and the model construction phase further builds machine learning models to better differentiate fake news and real news based on the feature representations [8].

Success has been achieved by implementing even quite simple artificial intelligence algorithms (such as naive Bayes classifier [16] [17], Logistic Regression [18] SVM [9] etc), neural networks algorithm (supervised or unsupervised learning approaches). These single algorithm solutions mostly are disadvantaged with the high false alarm rate problem leading to inaccurate classification. Based on some research works, it was observed that classification accuracy was significantly improved using more complex model [16] e.g. deep learning algorithms or using hybrid approaches [6] e.g. ensemble models. The researchers found that the hybrid algorithm performed slightly better than how they did individually. Despite these successes, it was noted that most existing hybrid fake news detection models are still unable to accurately classify news. The false positives

instances are usually high and true negatives instances are low.

From the literatures, it was established that the most effective approach to detect fake news based on news content and social content perspectives is by employing hybrid deep learning approaches. One of the recent and effective model was proposed in [24]. The model effectively minimized false positives as well as deliver better detection rates with an accuracy of 82% and intuitively identified relevant features associated with fake news stories without previous knowledge of the domain. However, the model being a supervised learning approach failed to construct a multitude decision at training time and also the variances used for classification of news was very high. These resulted to high false positive instances in the prediction.

Therefore, an improved fake news detection classification model that would incorporate both reinforcement and supervised approach is highly needed. Thus, in this paper, an attempt was made to develop an improved model with the capability of accurately detecting fake news on the social media platform by classifying news content as either real news or fake news using an hybrid approach of supervised and reinforcement learning approaches. This was with a view to assist the social media users in making decision to curb the dissemination of fake news on social media.

The rest of this paper is organized as follows. Section 2 covers the literature review while section 3 describes the methodology and presents the proposed architecture, section 4 discusses the results of the simulation. Finally, Section 5 concludes the study.

## II. RELATED WORKS

Several hybrid algorithms have been developed for fake news detection using data mining methods. In [19], a framework was proposed by synergizing AdaBoost and SVM classifiers on topological, crowdsourced and content-based (Hashtags, Mentions, URLs, and Phrases with sentiment extraction) features to detect political misinformation at the early stage on Twitter. The classifier performed well as a supervised learning approach, but the extracted features were substituted for content. This study showed that the only supervised learning approach is not adequate for accurate fake news detection. In [20], a hybrid deep model was proposed to address the timely problem of fake news detection by incorporating the text, the response an article receives, and the users who source it. The model was able to address the timely problem of fake news detection. Meanwhile, it was only based on user behaviour and does not incorporate reinforcement learning and humans in the learning process and thus affected the accuracy of the detection. In [21], the identification of fake news on Twitter was formulated as a binary machine learning problem. The result showed that instead of creating a small, but accurate hand-labeled dataset, using a large-scale dataset with inaccurate labels yields very good results. In [22], a tensor modelling of how to accurately

distinguish different categories of fake news based mainly on the content, where latent relations between articles and terms were captured as well as spatial or contextual relations between terms, towards unlocking the full potential of the content was proposed. The proposed algorithm, a supervised learning approach was able to identify different news categories with corpus with a percentage of 80%, but with high false positives rate. In [23], the tensor decomposition ensembles model was proposed by clustering the news content of fake news articles into different categories. They were able to do this by capturing spatial relations between terms for each article. This was able to identify all different categories of fake news within the dataset. Therefore, it was noted in [19][20][21][22] and [23] that they are mainly supervised learning approaches. Though they all yield a good detection rate, but they have not completely stopped the spread because of their inability to accurately classify news.

In [4], a stacked ensemble of 5 independent classifiers in the context of the natural language processing was developed to detect fake news. The stack ensemble was a two-layer classifier architecture that leverages predictions from weaker slave classifiers as features to a stronger master classifier. The classification experienced a difficulty in which the test set was far greater than that of the development data split which impacted results. Also in some cases, multiple layers of neural networks are required for effective training of models. This revealed that a stack ensemble method of classifiers is appropriate to detect fake news accurately. In [24], a framework that

detects and classifies fake news messages from Twitter posts using a hybrid of Convolutional Neural Networks and Long-Short Term Recurrent Neural Network models was proposed. The proposed work using this deep learning approach achieves 82% accuracy and intuitively identified relevant features associated with fake news stories without previous knowledge of the domain. The dataset used was limited because deep learning models such as CNN and RNN require much larger datasets. This framework revealed that accuracy of the fake news detection is better achieved using deep learning methods as compared to the conventional machine learning algorithms.

Despite all these encouraging results, there are still many open questions associated with fake news detection. One interesting direction would be to build models that will incorporate both unsupervised and reinforcement learning which could lead to more accurate and timely predictions.

### III. METHODOLOGY

In this study, an improvement was made on the fake news identification on Twitter network model proposed by [24] in which the synergy of Convolutional Neural Networks and Long-Short Term Recurrent Neural Network models was proposed to detect fake news on twitter. A PHEME dataset containing 5,800 tweets was used. The model conceptualisation and the detailed algorithm of the proposed model are presented as follows:

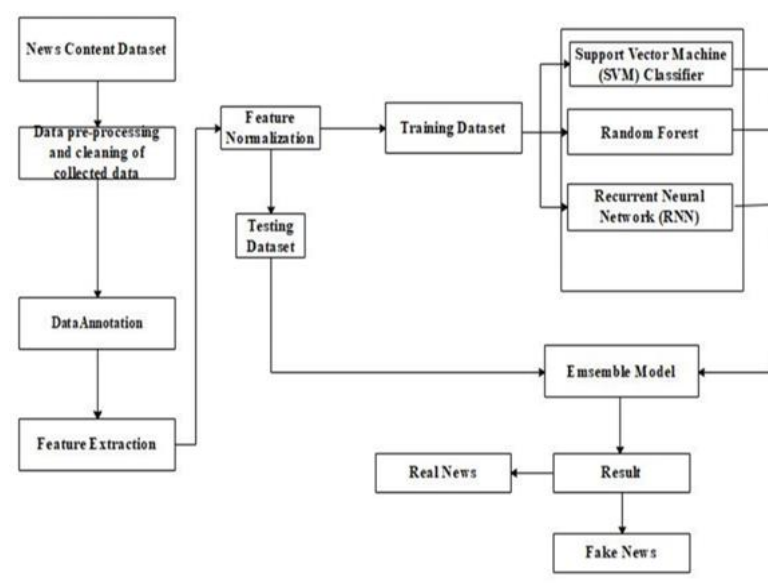


Fig.1. Proposed Model Block Diagram

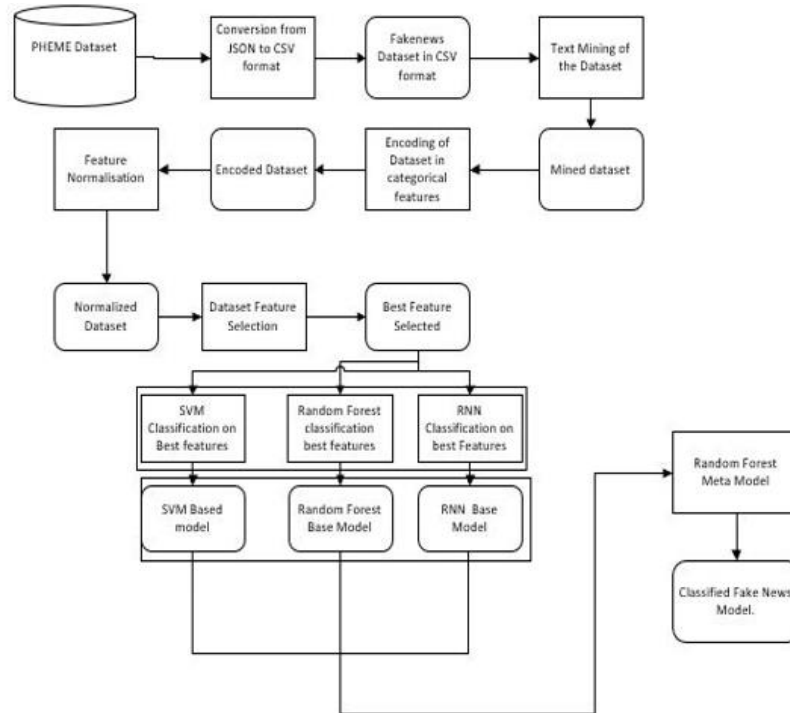


Fig.2. Block Diagram of Entire Fake News Classification Process

### A. Conceptual Model Description

This proposed model uses a stacking ensemble method of three machine learning algorithms namely: Support Vector Machines (SVM), Recurrent Neural Networks (RNN) and Random Forest which are supervised machine learning algorithms as against the existing model which made use of a hybrid of Convolutional Neural Network and Recurrent Neural Networks algorithms. This method was adopted to improve the prediction rate of the classification model which in turn improves the ability of the model to accurately classify news as either fake or real news. The proposed conceptual block diagram for the classification model for fake news detection on social media is shown in Fig. 1 and the block diagram for the entire classification process is shown in Fig. 2. These Figures shows the details of the data collection and pre-processing phases involved in the model. This proposed model uses a stacking ensemble method of machine learning and deep learning algorithms in order to improve the prediction rate of the classification model which in turn improves the ability of the model to accurately classify news as either fake or real news.

### B. Model Development Process

The stages applied to the development of the model is as follows:

1. Dataset used was obtained from PHEME fake news dataset and analysed to know the true state of the dataset, it was cleaned by removing special characters, html tags and other unclean texts.
2. Tf-idf was used to convert the cleaned dataset into vector. The labelled features which contains real news and fake news texts was converted to

numerical format using categorical encoding algorithm as shown in Fig. 3. This algorithm was employed to handle this since ensemble algorithms cannot process categorical variables and the performance of the proposed model depends on the function of encoded variables.

3. The dataset features were then normalized to ensure that the features are in the same scale to enhance the performance of the model
4. Relevant features were selected from the normalized dataset using entropy-based feature selection process. The selected features were then fitted into Support Vector Machine, Random Forest, and Recurrent Neural Network learning algorithms separately to generate the base models. The base models were fitted into the meta-model to generate the final model which is the stack ensemble model
5. The generated stack ensemble model now classifies the news instances into real or fake. The model was then tested with new network instances and the classification generated serves as the result of the model.
6. The performance of the model was then evaluated by benchmarking with the existing classification model [19] using detection accuracy, sensitivity, false alarm rate and precision as performance metrics.

#### a. Data Pre-processing

The PHEME dataset which contains five major news events was stored in .csv format. Pandas was used to combine all the news events, it has a total of 103212 news instances which is the combination of original tweets and user reaction to the tweets.

The dataset has 34 features including the date the news event was created. The features are; coordinates created, favorite\_count, has\_url, hashtags\_count, id, in\_reply\_id, in\_reply\_user, is\_rumor, symbols\_count, text, etc. The text features and target variable (is\_rumour) was used cleaned to be used for the classification model.

#### b. Feature Selection

The dataset records used for this study is a multi-dimensional dataset containing features that are not relevant in formulating the proposed model. Thus, the need to do dimensionality reduction by selecting the best features from the high dimensional dataset. Mutual information gain features selection algorithm was employed in selecting the best features from the dataset based on the information gain.

#### c. Text to Vector Conversion

The bag of words generated after cleaning the dataset was further processed by converting to vectors using the Term Frequency- Inverted Document Frequency (tf-idf) tool as one of the texts to vector conversion tool in Python. This is an important step in developing a machine learning/deep learning model which will allow it to understand the dataset. The term frequency is the type of path that appropriates the counts of words present in the document and figures out the inequality between the documents. Each document is characterized by a vector that contains the word count. Each word count is transformed into the possibility of such word current in the documents. The idf is a statistical measure used in this study to evaluate how important a word is to a document in the dataset. It was used to measure the frequency of the time in the collection for weighting and ranking for words in the bag of words. Tf-idf as used in this study measure the sum of term frequency like number multiplied by term importance.

#### d. Feature Normalization

The Vectorized dataset was normalized by giving all features an equal weight in other to decrease computation time and increase the detection accuracy of the proposed model in a pre-processing dataset for the machine learning model. The features are scaled into a specific range of 0 and 1 in other to ensure that all features are in the same scale. Min-Max normalization was used to scale the features from a range of values to a new range of values.

### C. Model Formulation

The detailed algorithm of the proposed model formulation is as presented in Fig. 4. The formulation started with the identification of the spread of fake news as a problem on social media and ended with predictions of either fake or real news. The classification algorithm used to categorize each data into the class (Fake and real news) to identify the category or class to which each data will fall under is presented in Fig. 5. The algorithm of the stacking ensemble model, which revealed how the predictions from the individual algorithms are used as

features to build the proposed model is also presented in Fig. 6. This algorithm averages out noise from the classifiers to enable improved predictive accuracy performance of the model. The Support Vector Machine (SVM) algorithm, a supervised learning algorithm used for classification analysis is shown in Fig. 7. The main reason for using SVM is because the problem of fake news detection is not linearly separable. SVM was used to find the best separating hyperplane also called decision boundary. Also, to design a hyperplane that classifies all dataset into two classes i.e. training and testing dataset. SVM helps to label the datasets appropriately since mislabelling can decrease model performance. Random Forest which was used as the reinforcement learning method of the ensemble model operates by constructing a multitude decision at training time and also used to reduce the variances used for classification of news. The algorithm is presented in Fig. 8. By incorporating the reinforcement learning into the modelling process will enhance the detection rate of the model. The RNN algorithm was employed to model the sequence of the data in order to remember the all the inputs throughout the time of modelling. This helps to address the problem of overfitting, underfitting and leakage of the training dataset, to improve the prediction accuracy of the model. Thus, the pre-processed dataset was used to train each of the algorithms individually as base-level classifiers. The predictions of the three models are then used as inputs in a Random Forest algorithm (meta-classifier) to form the final ensemble model.

```

Input: X(Dataset)
Output: X, Identical Attributes of X
1: begin
2:   func textToNumericConverter(df):
3:     [row, column]=columnvalues(X)
4:     for each column in columns (X) do
5:       tex digit vals = {}
6:       func convertToInt(val):
7:         return tex digit vals[val]
8:       if col(type)6= (num) and col(type)6= (float): then
9:         df[col]contents = df[col].values.tolist()
10:        uniqueelements = set(column contents)
11:        x=0
12:        for unique in unique elements do
13:          if unique not in tex digit vals then
14:            tex digit vals[unique] = x
15:            x +=1
16:          end if
17:          df[col] = list(map(convertToInt, df[col]))
18:        end for
19:      end if
20:    return df
21:  end for
22:  df = textToNumericConverter(df)
23: end

```

Fig.3. Categorical Attribute Encoding Algorithm

1. Identification of fake news spread on Twitter.
2. Select a data source containing news contents
3. Data preprocessing, cleaning and annotation
  - Conversion of the dataset from JavaScript Object Notation (.json) to Comma Separated Value (.csv) format comprising of the user tweets and user reactions to the tweets.
  - Load the combined dataset into data frame in python using pandas
  - Extraction of text and target variable (is\_rumour) columns from the dataset.
  - Removal of HTML tags, characters and symbol.
  - Splitting of sentences in the text into individual words.
  - Removal of stop words from the text.
  - Replacing variation of words with the root words.
  - Converting text in uppercase to lowercase.
  - Creating bag of words from the cleaned dataset.
  - Encoding of string characters into numerical values using the categorical attribute encoding algorithm (Algorithm 1)
4. Feature Selection and Normalisation
  - Conversion of the bag of words into vectors using tf-idf
 
$$W_{t,d} = (1 + \log(1 + \text{tf}_{t,d})) \cdot \log_{10} N / \text{df}_t$$
  - Normalization of the vectors created from the bag of words using Min-Max normalization:
 
$$Q = \left( \frac{p - \text{Min}(p)}{\text{Max}(p) - \text{Min}(p)} \right) * (N - M) + M$$

where the value of P feature needs to be normalized into value Q.  $\text{Min}(P)$  and  $\text{Max}(P)$  is the minimum and maximum values of feature P respectively.  $M$  and  $N$  indicates Lower and Upper Values respectively in the new range. (0,1) is used to normalized the features of  $P$ , this make  $Q$  to be in the range 0 and 1

    - Dividing the dataset into two i.e. Testing and Training dataset
    - Calculation of the relevance of each feature to the label feature using Entropy Based (Mutual Information Gain) feature selection.
    - Select Best Feature with Highest information gain using algorithm 3
5. Building the base models by applying Support Vector Machine (SVM), Random Forest (RF) and Recurrent Network (RNN) algorithms separately on the training dataset.
6. Applying the test dataset on the base models separately.
7. Stacking of the base models to form the meta model
8. Application of the smaller part of the dataset divided for testing the fake news stack model.

Fig.4. Proposed Model for fake News Detection

```

Input: Test datasets ( $M_i = 1, \dots, n$ ), Meta model
Output: Classified Test dataset
1. begin
2. Input Test dataset into meta model
3. for each Test dataset  $M_i$  do
   Classify Test dataset class as is_rumour or rumour
4. end for
5. Return classified Test dataset from the meta model
6. end
    
```

Fig.5. Classifying Test Dataset

```

Input: training data  $D = \{x_i, y_i\}$ 
Output: ensemble classifier H
Step 1: learn base-level classifiers
1. for  $t = 1$  to  $T$  do
   Learn  $h_t$  based on  $D$ 
2. end for
Step 2: Construct new data set of predictions
3. for  $i = 1$  to  $m$  do
    $D_h = \{x_i, y_i\}$ , where  $x_i = \{h_1(x_i), \dots, h_T(x_i)\}$ 
4. end for
Step 3: Learn a meta-classifier
5. learn H based on  $D_h$ 
6. return H
    
```

Fig.6. Stacking Ensemble Algorithm

```

Input: X: Training set.  $\delta$ : Threshold.
Output:  $X_R : X_R \subset X_{sk}$ 
 $|X_R| \ll |X|$ 
1. Begin Train a decision tree  $T // X_R$ 
   Begins empty  $X_R \leftarrow NULL$ 
2. For each leaf  $\mathcal{L}_i$  of  $T$  do
3. for each opposite class neighbour  $\mathcal{L}_j$  do
4. If entropy of  $\mathcal{L}_j$  is low then
5. Use  $\mathcal{L}_i$  and  $\mathcal{L}_j$  to build  $X^*$ ;
6. Compute  $\omega$ 
7. Add  $x_j \in \mathcal{L}_j$  to  $X_R$  according to (12);
8. end for
9. else
10.  $X_R \leftarrow X_R \cup \mathcal{L}_j$ ;
11. end if
12. end for
    
```

Fig.7. Support Vector Machine (SVM)

```

OUTPUT: Random Forest fake news classification model
BEGIN
1. For each tree in the forest do
2. Sample N datapoints with replacement;
   For each internal node in the tree do
       Randomly sample  $m$  attributes ( $m = \sqrt{d}$ )
       Select the feature from this randomly selected set
       that provides the most information gain
   end For
3. Return Fully grown decision tree
4. end
5. return Random Forest
END
    
```

Fig.8. Random Forest Classification

#### IV. RESULTS AND DISCUSSIONS

The proposed model was simulated and evaluated using Python 3.7.3 programming language with its machine learning library (i.e. Scikit-learn library). Jupyter notebook was used for the coding while Panda, a data analysis library was used for the pre-processing of the dataset. The detailed results are presented as follows.

A. Data Pre-processing Results

The PHEME Dataset contains source tweets and user reactions to the tweets have a total number of 103,212 news instances which has been classified into 31,430 fake news instances and 71,782 real news instances as shown in Fig. 9. The graphical representation of the proportion of source tweets to the user reaction of each news event is also presented in Fig. 10. The overall dataset was divided into eighty percent as training dataset and twenty percent as test dataset which are 82,569 and 20,642 respectively.

For validation of the dataset, cross- validation technique was applied using the *train\_test\_split* of the scikit-learn library. The eighty percent (82,569) of the dataset was further splitted into 66,055 (80%) and 16,514 (20%). The 66,055 news instances were used to train the base classifiers and 16,514 news instances were used to test the base classifiers. The remaining news instances in the dataset (20,642) was used to test the stack ensemble model.

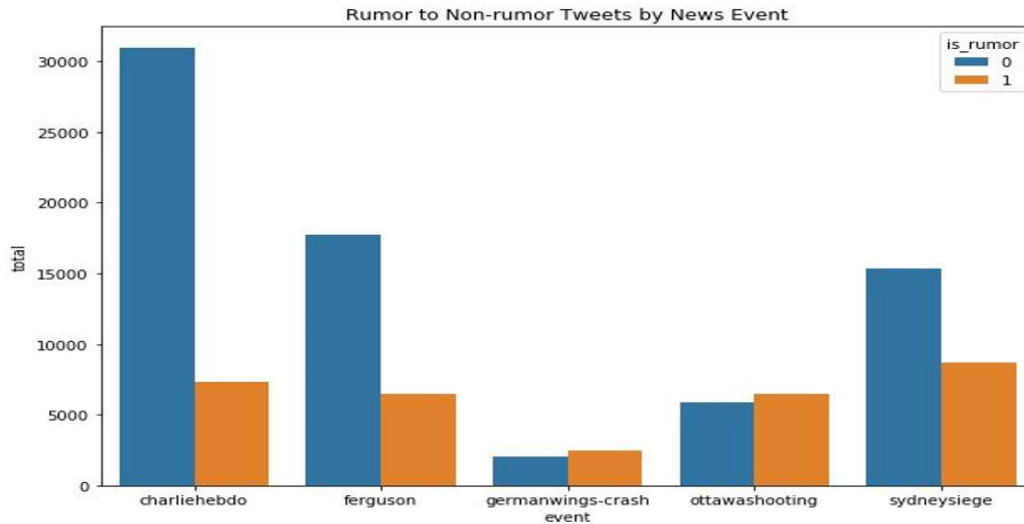


Fig.9. Bar Chart Showing the Proportion of Fake News to Real News.

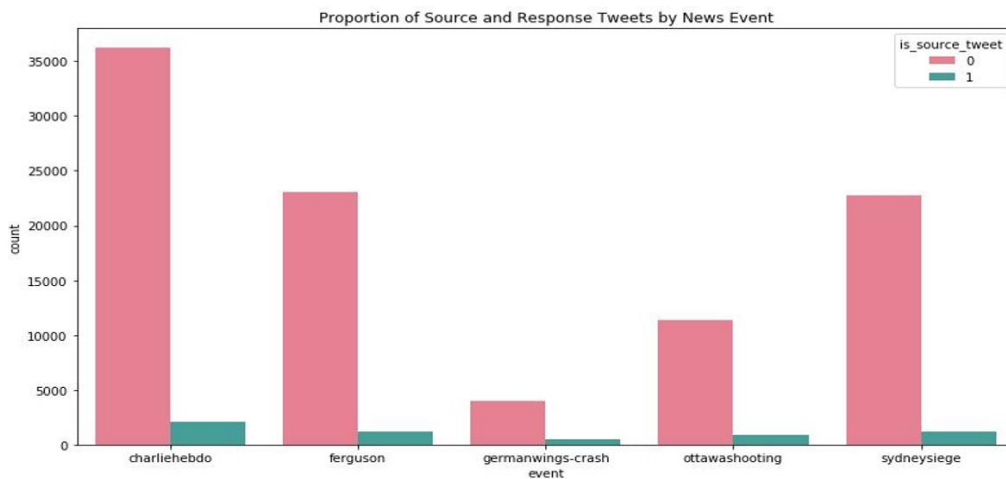


Fig. 10. Bar chart showing the proportion of Source tweets and User-reactions

B. Simulation Result

The training dataset which is 66,055 (60% of the PHEME dataset) was used on each of the base learner classifiers. Twenty percent (20%) of the dataset which is 16,514 was used to test the base learner classifiers. The output of predictions of the base models was combined and used as input to train the Meta classifier which is Random Forest to get the final prediction of the stacking ensemble model. After developing the stack ensemble model, the remaining news instances in the dataset

(20,642) was used to test the performance of the stack ensemble model.

The result from the confusion matrix in Table 1 showed that the proposed model was able to correctly classify 12,544 (87% of real news instances) as real news (i.e. True Positives (TP)), while 4329 (70% of fake news instances) was correctly classified as fake news instances (i.e. True Negatives (TN)). The result also showed that 1,885 (13% of real news instances) were misclassified as fake news instances (i.e. False Positives (FP)), while 1,885 (30% fake news instances) were misclassified as

real news instances (i.e. False Negatives (FN)). This shows that the proposed model detects fake news with a reduced false alarm rate.

The testing dataset was also fed into the existing model, the result from the confusion matrix as shown in Table 2, shows that 14,429 (100% of real news instances) were classified as real news. The existing model confusion matrix result also shows that the model was unable to detect any fake news instances as fake news, it then classified all 6,214 fake news instances as real news (i.e. False Positives), then, none of the real news instances were misclassified, this makes the existing model to predict with high rate false alarm.

Table 1. Confusion Matrix Result of the Proposed Model

	Real News (Predicted)	Fake News (Predicted)
Actual (Real News)	12544(TP)	1885(FN)
Actual (Fake News)	1885(FP)	4329(TN)

Table 2. Confusion Matrix Result for the Existing Model

	Real News (Predicted)	Fake News (Predicted)
Actual (Real News)	14429(TP)	0(FN)
Actual (Fake News)	6214(FP)	0(TN)

The result of the study showed that the stack-ensemble of SVM, Random forest and Recurrent Neural Network (base learners) have accuracy of 69.2%, 71.9% and 86.6% respectively while, Random Forest (meta-classifier) trained with the output of the predictions of the base learners has an accuracy and sensitivity of 81.9% and 82% respectively which is better than the existing model which has a percentage of 70%. The existing model was unable to detect any labelled fake news instances as fake news while the developed model recorded a precision of 82%. The simulation result showed that the developed model could detect more fake news instances than the existing model.

C. Model Evaluation Results

The performance of the proposed model was evaluated by benchmarking with the existing model [19] using detection accuracy, sensitivity, false alarm rate and precision as performance metrics according to [8]. All these performance metrics used were calculated using the result of the confusion matrix table which is categorized under true negative, true positive, false positive and false negative respectively as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} * 100 \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \left( \frac{TP}{TP + FP} \right) * 100 \tag{3}$$

The result of the evaluation and their percentage difference is presented in Table 3. The table shows that the proposed model has a 17.25% higher detection accuracy and 15.78% sensitivity, but a lesser 22.2% precision than the existing model. The higher precision of the existing model emanated because the existing model was unable to detect any fake news instances as fake news, which led to the high positive rates.

Tables 4 and 5 show the screenshot of the classification for the proposed model and the existing model indicating precision, sensitivity, and f1-score. Fig. 11 showed the Area under Curve of 88% for the proposed model. Figs. 12 13 and 14 are the graphs for the precision and sensitivity that show the benchmarking results of the existing model and the proposed model respectively.

The evaluation results indicate that the proposed model is more sensitive and accurate to classify news as either fake or real than the existing model.

Table 3. Percentage Difference between the Existing Model and Proposed Model

S/N	Metrics	Existing model	Proposed model	Percentage difference
1	Accuracy	69.89%	81.9%	17.25%
2	Precision	100%	82%	-22.2%
3	Sensitivity	70%	82%	15.78%

Table 4. Classification Table for Proposed Model

	Precision	recall	F1-score	Support
0	0.87	0.87	0.87	14429
1	0.70	0.70	0.70	6214
micro avg	0.82	0.82	0.82	20643
macro avg	0.78	0.78	0.78	20643
weighted avg	0.82	0.82	0.82	20643

Table 5. Classification Table for Existing Model

	Precision	recall	F1-score	Support
0	1.00	0.70	0.82	20643
1	0.00	0.00	0.00	0
micro avg	0.70	0.70	0.70	20643
macro avg	0.50	0.35	0.41	20643
weighted avg	1.00	0.70	0.82	20643

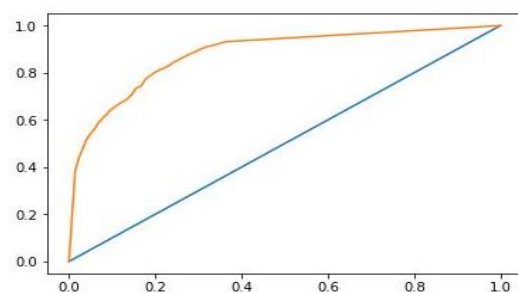


Fig.11. Area under Curve of the proposed model



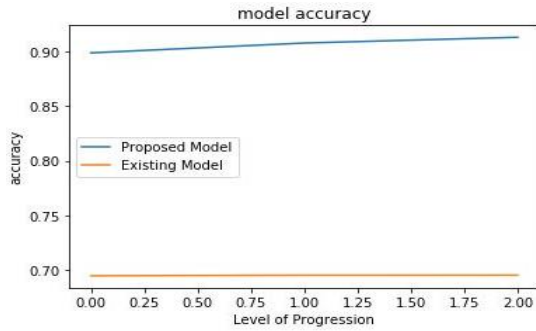


Fig.12. Detection Accuracy of the Existing and Proposed Model

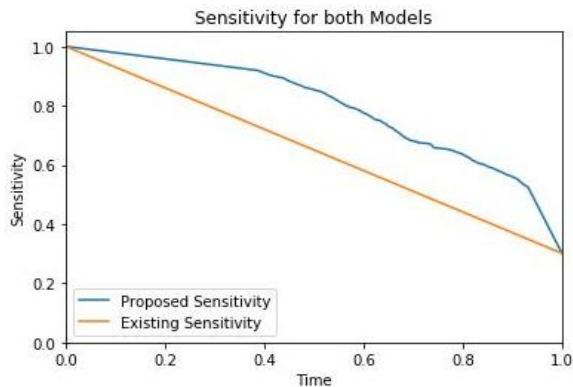


Fig.13. Sensitivity of the Existing and Proposed Model.

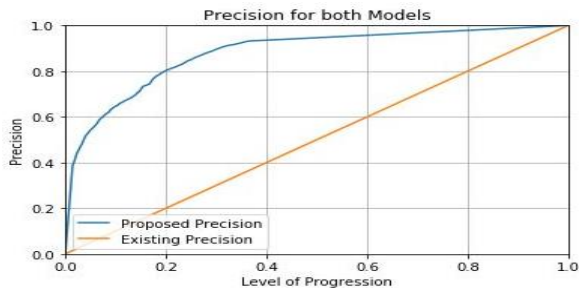


Fig.14. Precision of the Existing and Proposed Model.

## V. CONCLUSION

In this study, a classification model for fake news detection on social media using a stack ensemble method of three algorithms to increase the detection accuracy was developed. The proposed model was simulated and it was obvious in the results that the proposed model has a high reduction in the instances of false positives states and high detection accuracy and sensitivity rates, thus has a better performance in detecting fake news instances accurately. Therefore, the proposed model can be adapted by the Social media administrators to combat the spread of fake news. The future works is directed to having a model that will address the timely prediction of the fake news in social media.

## REFERENCES

[1] A. Perrin, Social Media Usage: 2005-2015. Washington, D.C.: Pew Internet & American Life Project. Retrieved October 12, 2015 from

<http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/>, 2015.

- [2] J. Clement, Number of internet users in Nigeria from 2017 to 2023. Available at <https://www.statista.com/statistics/183849/internet-users-nigeria/>, 2019.
- [3] Africa Practice. "Social Media Landscape in Nigeria." Last modified 2014. Accessed 16 June 2016. <http://www.africappractice.com/wp-content/uploads/2014/04/Africa-PracticeSocial-Media-Landscape-Vol-1.pdf>, 2014.
- [4] J. Thorne, M. Chen, G. Myrianthous, J. Pu., X. Wang, A. Vlachos, Fake news stance detection using stacked ensemble of classifiers. Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, 2017, pp 80–83, DOI: 10.18653/v1/W17-4214.
- [5] S. Schifferes, N. Newman, N. Thurman, D. Corney, A. G'oker, C. Martin, Identifying and verifying news through social media: Developing a user-centred tool for professional journalists. Digital journalism, 2014, 2(3):406–418.
- [6] K. Stahl, Fake news detection in social media. Available at [https://www.csustan.edu/sites/default/files/groups/University%20Honors%20Program/Journals/02\\_stahl.pdf](https://www.csustan.edu/sites/default/files/groups/University%20Honors%20Program/Journals/02_stahl.pdf), 2018.
- [7] A. Ceron, L. Curini, S. M. Iacus, G. Porro, Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to Italy and France. New media & society, 2014, 16(2):340–358.
- [8] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter. June 2017, 19(1): 22-36, 10.1145/3137597.3137600.
- [9] B. D. Horne, S. Adali, This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In proceedings of the Eleventh International AAAI Conference on Web and Social Media, 2017.
- [10] C. Silverman, Here are 50 of the biggest fake news hits on Facebook from 2016. BuzzFeed, <https://www.buzzfeed.com/craigsilverman/top-fake-news-of-2016>, 2016.
- [11] E. Ferrara, Manipulation and abuse on social media by emilio ferrara with ching-man au yeung as coordinator. ACM SIGWEB Newsletter Spring, 2015, 4.
- [12] J Keller, A fake AP tweet sinks the DOW for an instant. Bloomberg Businessweek, 2013.
- [13] F. Álvaro, O. Luciana, The current state of fake news: challenges and opportunities. Procedia Computer Science, 2017, 121:817-825, DOI: 10.1016/j.procs.2017.11.106.
- [14] X. Zhou, R. Zafarani, Fake News: A Survey of Research, Detection Methods, and Opportunities. ACM Comput. Surv, 2018, 1:1- 40.
- [15] Y. Wu, P. K. Agarwal, C. Li, J. Yang. Toward computational fact-checking. Proceedings of the VLDB Endowment, 2014, 7(7):589-600
- [16] M. Granik, V. Mesyura, Fake News Detection Using Naive Bayes Classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017: 900-903.
- [17] V. Qazvinian, E. Rosengren, D. R. Radev, Q. Mei, Rumor has it: Identifying misinformation in microblogs. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011:1589–1599.
- [18] W. Ferreira and A. Vlachos, Emergent: a novel data-set

for stance classification. In proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016:1163–1168, DOI:10.18653/v1/N16-1138

- [19] J. Ratkiewicz, M. Conover, M. R. Meiss, B. Goncalves, A. Flammini, F. Menczer, Detecting and tracking political abuse in social media. ICWSM, 2011, 11:297–304.
- [20] N. Ruchansky, S. Seo, Y. Liu, CSI: A Hybrid Deep Model for Fake News Detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17), 2017:797-806. DOI: 10.1145/3132847.3132877
- [21] S. Helmstetter, H. Paulheim, Weakly Supervised Learning for Fake News Detection on Twitter. In proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018: 274-279
- [22] E. E Papalexakis, C. Faloutsos, N. D Sidiropoulos, Tensors for data mining and data fusion: Models, applications, and scalable algorithms. ACM Transactions on Intelligent Systems and Technology (TIST), 2017, 8 (2): 16.
- [23] S. Hosseinimotlagh, E. E. Papalexakis, Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. Misinformation and Misbehavior Mining on the Web Workshop held in conjunction with WSDM , 2018.
- [24] O. Ajao D. Bhowmik, S. Zargari, Fake news identification on twitter with hybrid cnn and rnn models. In Proceedings of the 9th International Conference on Social Media and Society, 2018: 226–230.
- Journal of Information Technology and Computer Science(IJITCS), Vol.12, No.1, pp.34-43, 2020. DOI: 10.5815/ijitcs.2020.01.05

### Authors' Profiles



**Bodunde AKINYEMI** is a Senior Lecturer at Obafemi Awolowo University, Ile –Ife. Her current research interest includes CyberSecurity, Data communication and networking. She is a Member of the Nigeria Computer Society (MNCS) and a chartered IT practitioner (MCPN).



**Oluwakemi ADEWUSI** is a post-graduate student of Obafemi Awolowo University, Ile-Ife, Nigeria. His current research interest includes cyber security and Data mining.



**Adedoyin OYEBADE** is a post-graduate student of Obafemi Awolowo University, Ile-Ife, Nigeria. His current research interest includes cyber security and Data mining.

**How to cite this paper:** Bodunde Akinyemi, Oluwakemi Adewusi, Adedoyin Oyebade, "An Improved Classification Model for Fake News Detection in Social Media", International