International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 07, Volume 8 (July 2021) https://www.ijirae.com/archives

DEVELOPMENT OF A CONCEPTUAL FRAMEWORK AND A MEASUREMENT MODEL FOR THE DETECTION OF FAKE NEWS

Samuel A. Oyeniyi*
Cyber Security Science Federal University of Technology, Minna, Nigeria
Sammykul01@gmail.com
Joseph A. Ojeniyi
Cyber Security Science Federal University of Technology, Minna, Nigeria
ojeniyija@futminna.edu.ng

Publication History
Manuscript Reference No: **IJIRAE/RS/Vol.08/Issue07/JLAE10080**
Research Article | Open Access
Peer-review: Double-blind Peer-reviewed
Article ID: IJIRAE/RS/Vol.08/Issue07/JLAE10080
Received: 02, July 2021
Accepted: 12, July 2021
Published Online: 30, July 2021
Volume 2021 | Article ID JLAE10080 | https://doi.org/10.26562/ijirae.2021.v0807.001
Samuel,Joseph (2021). Development of a Conceptual Framework and a measurement model for the detection of fake news. IJIRAE:: International Journal of Innovative Research in Advanced Engineering, Vol: VIII,138-147
doi: https://doi.org/10.26562/ijirae.2021.v0807.001
Editor-Chief: Dr.A.Arul Lawrence Selvakumar, Chief Editor, IJIRAE, AM Publications, India

Abstract: Fake news has been there since before the advent of the Internet. It has had an immense impact on our modern society. Detecting fake news is an important step. Although there are various ways and methods in which fake news can be detected and solved. In this research paper we discuss the various conceptual frameworks and how they affect fake news. It further shows the development of the conceptual framework and the measurement model used; showing which of the frameworks fake news is most likely to surface through. The objective of the research is to design a conceptual framework for fake news detection, whereby developing measurement model for fake news detection, and the framework and model are evaluated for fake news detection. Fake news detection approaches can be divided as: creator and user features, news content features and social context features. A survey was taken based on this feature via questionnaire to determine in which feature, fake news can be quickly spotted. Results: Results shows that fake news can be easily spotted in the creator and user feature, this feature was then used to perform a feature selection on a fake news dataset which gave better accuracy.

Keywords: development model measurement framework conceptual

I. INTRODUCTION

Fake news or junk news or pseudo-news is a type of yellow journalism or propaganda that consists of deliberate disinformation or hoaxes spread via traditional print and broadcast news media or online social media. The false information is often caused by reporters paying sources for stories, an unethical practice called check book journalism. Digital news has brought back and increased the usage of fake news, or yellow journalism. The news is then often reverberated as misinformation in social media but occasionally finds its way to the mainstream media as well. Fake news is a longstanding problem that has affected all types of media: printed media, radio, television and recently digital social media. The "Great Moon Hoax"1 in 1835 is known as one of the earliest examples of fake news, in which the New York Sun published a series of articles about the supposed discovery of life on the moon. Social media is an environment that enables the rapid productions and dissemination of information at a very low cost. Due to its massive dissemination capabilities, digital and social media can reach out to millions of users within minutes. With the increase in popularity, social media has become the main source of information for many people worldwide. Despite these advantages, social media is considered to be the news production media which varies a lot from the traditional news media. Consequently, the quality of information produced by them is considered to be lower than the traditional news media. In digital media, the boundary between news production and information creation is gradually blurring. Due to the low quality of news, there is a need to permanently asses the quality of news published in the social media [11]. Fake news has become increasingly prevalent over the last few years, with over 100 incorrect articles and rumours spread incessantly just with regard to the 2016 United States presidential election.

These fake news articles tend to come from satirical news websites or individual websites with an incentive to propagate false information, either as click bait or to serve a purpose. Since they typically hope to intentionally promote incorrect information, such articles are quite difficult to detect. In the research [1] made use of a tool that can identify and remove fake sites from the results provided to a user by a search engine or a social media news feed. And also [7] showed us in their research that even quite simple artificial intelligence algorithm (such as naive Bayes classifier) may show a good result on such an important problem as fake news classification. Experimental results show that machine learning models combined with the proposed data pre-processing method outperform baselines [9][10] Shows Experimental result on large miscellaneous events dataset demonstrates the effectiveness of the proposed approach in identifying fake tweets. [12] Wrote a paper focusing on distinguishing satire or parody and fabricated content using the Fake vs. Satire public dataset by reviewing existing literature in two phases: characterization and detection. Detecting fake news on social media poses several new and challenging research problems. Though fake news itself is not a new problem nation or groups have been using the news media to execute propaganda or influence operations for centuries–the rise of web-generated news on social media makes fake news a more powerful force that challenges traditional journalistic norms. There are several characteristics of this problem that make it uniquely challenging for automated detection. First, fake news is intentionally writ- ten to mislead readers, which makes it nontrivial to detect simply based on news content. The content of fake news is rather diverse in terms of topics, styles and media platforms, and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. For example, fake news may cite true evidence within the in- correct context to support a non-factual claim.

II. METHODOLOGY

A. Research Design

This research will adopt a semi structured survey design. The survey research method will have two identifying features. First, it is based on the question asked in the questionnaires which are gotten from the identified framework. Second, the data are collected by having each individual complete a questionnaire. The researcher will obtain data from the respondents by means of questionnaire which is distributed using online based on Google forms.

1) Problem Formulation: As well known, News can be said to be an information that is published in newspapers, broadcasted on radios and televisions and also mostly on social media platforms about recent events in a country or world or in a particular area of activity. In as we have news there is a thin line between news and fake news. Fake news is said to be a type of propaganda that exist of deliberate misinformation spread via traditional print and broadcast news media or online social media platforms, in this research we are making use of a fake news conceptual framework and using a measurement model to determine where or which area fake news can detected.

2) Population Sampling Techniques: The type of sampling technique to be used for this research work is the Simple Random Sampling. Simple Random Sampling (SRS) is a subset of individuals chosen from a larger set of population. Each individual are chosen by chance from various locations and geographical zones.

3) Reliability of Research Instrument: Undergoing this research various instrument were used in other to obtain an accurate and positive result. In the process of this research Google forms was used for the creation and distribution of the questions and questionnaires respectively which has a high success rate of use with less failure rate, which made it the ideal tool for the research, not forgetting easy access for distribution on a larger scale to various geographical zones. And also the use of Statistical Package for the Social Sciences (SPSS) software.

4) Mode of Data Collection: Data collected with the instrument (questionnaire) via the use of Google forms were listed out, where the responses were gotten on a spreadsheet. The quantitative data were analyzed using Statistical Package for the Social Sciences (SPSS) software. Statistical means such as frequency tables were used to reduce the raw data into manageable proportions.

B. Architectural Design

The architectural aspect of this research depicts the sources of news that encompasses of creator/user based features, news content feature and social context based features which is linked to the distribution which serves as the questionnaires given to the users, and the response is then analyzed as obtained from the distribution which leads to the detection

C. Development of Conceptual Framework for Detecting Fake News.

1) Conceptual Framework: At the review on the stat-of-the-art studies on fake news detection. Overall categorizations of the current research on online fake news detection, from which we can discover the differences of detecting different types of false information in terms of features, data mining algorithms, and platforms. Online social media has indeed rise the widespread dissemination of online fake news. Most information distributed via social networks are huge, quick, lengthy, diverse, and heterogeneous, noting that online false information can cause severe impact to the whole society. As a result, more research are undergoing on detecting false information and fake news on online social media, the number of studies focusing on fake news detection is more significant than other topics (e.g., rumour or satire detection). There are mainly two types of systems: practical-based approaches and research-based approaches.

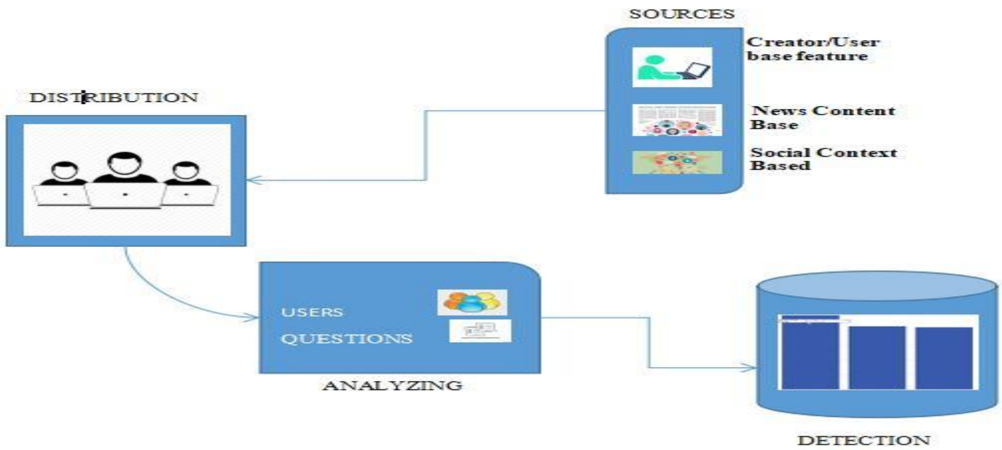

Fig 1. Architectural Design

Practical-based approaches are the verification of a piece of news can not only depend on the news content, creator of the news and the social context of the news are also influential factors.With the additional information such as the credibility of the news creator, and the underlying distribution pattern of the news, we can have better understanding of the news, and make more accurate prediction. There are three types of models for research-based approaches: component-based category, data mining based category, and implement-based category. And each model and its subcategories are shortly discussed as follows.

2) Component-based category: fake news contains four major components: Fake news creator, target victim, fake news content and fake news social context. Based on the analysis of different components, fake news detection approaches can be divided as: creator and user analysis, news content analysis and social context analysis.

3) Creator and user analysis: There are extensive attempt and efforts on the analysis of malicious accounts on social media. Being exposed to a large amount of unproven messages, online users lack the clues to evaluate the credibility of the social information. Malicious social media accounts intent to manipulate people's decision and pollute the truth news content by purposely spreading misinformation, so creator and user analysis is a critical aspect for fake news detection. Creator and user analysis can be categorized across the following differences: user profiling analysis, temporal and posting behaviour analysis, credibility-related analysis, and sentiment-related analysis.

4) User profiling analysis: The basic user profiling information includes the language used by the account, the geographic locations of the account, the account creation time, if the account is verified or not, how many posts/tweets does the account have.

5) Temporal and posting behaviour analysis*:* Temporal behaviour reveal the temporal patterns of the online social account, such as the signal similarity to a Poisson process the average time between two consecutive posts, the frequency of replying, sharing, mentioning, and so on.

6) Credibility-related information: The numbers of friends and followers are also good features for differentiating malicious accounts and legitimate users.

7) Sentiment-related analysis: Sentiment-related factors are also key attributes for suspicious account identification. By triggering anomalous emotional response, malicious accounts can exaggerate the facts and mislead legitimate users.

8) News Content analysis: fake news usually contains the physical content (like title, body text, image or video), and the non-physical content (like purpose, sentiment, and news topics).

9) Social context analysis: social context is the social environment in which the news is been spread. Social context analysis is the study of how fast and broad the social data is spread out, and how online users interact with each other. However, most of the recent approaches for online fake news detection are related to direct news content analysis.

10) Creator/User based features: Creator/User-based highlights have been generally utilized for suspicious online record recognition, these highlights expect to catch the remarkable attributes of suspicious client accounts or non-human accounts, and can be sorted as user profiling features, user credibility feature and behaviour based feature.

11) User profiling feature: User profiling feature incorporate the fundamental user data, for example, account name, geolocation in-development, the information of enlistment of the client, confirmed or not, has depiction or not, etc.

12) User credibility feature: User credibility feature record the effect and the credibility of the online record, incorporate the validity score of the user, the quantity of companions and supporters of the user, the proportion between the client's companions and devotees, the complete number of tweets/posts of the user.

13) User behaviour features: User behaviour highlights can be considered as a major aspect of social setting list of capabilities, which we may examine in the accompanying. User conduct highlights expect to acquire user standard of conduct for both tricky users and genuine users.

14) News content-based feature: News content-based features are unequivocal pieces of information for fake news identification, and they are most usually utilized properties for counterfeit news portrayal and recognition investigation. They can be ordered as linguistic and syntactic-based features, style-based features and visual-based features.

15) Linguistic and Syntactic-based features: Linguistic and syntactic-based features allude to the crucial part, structure and semantics for common language. Albeit counterfeit news content are constantly produced purposefully for misdirecting on the web users, linguistic and syntactic-based features are as yet significant sources for suspicious news examination.

16) Style-based features: Style-based features intend to uncover the various qualities of composing styles for fake news creators. Albeit more often than not, fake news writers attempt to emulate the composing style of an normal news authors to deceive the online readers, there are still a few contrasts which can used to segregate fake news creators and genuine news creators.

17) Visual-based features: The pictures or recordings contained in content are basic signs for distinguishing suspicious or beguiling data. Late studies investigate visual-based features for online deception distinguishing proof.

18) Social context-based features: Social context-based features are intended to mirror the distribution example of the online news, and the cooperation between online users. And they can be outlined into the accompanying three sorts: 2network-based features, distribution-based features, and temporal-based features.

19) Network-based features: Network-based investigation plan to concentrate on a gathering of comparable online users, in term of alternate points of view, similar to area, training foundation, and propensities. And network-based features are chosen and separated based on explicit networks and can be utilized to examine the novel attributes of specific networks, and the likeness and disparity of various online records.

20) Distribution-based features: Distribution-based features can catch the unmistakable dissemination example of online news. Typically a proliferation tree can be worked to encourage describing the distribution idea of a bit of news.

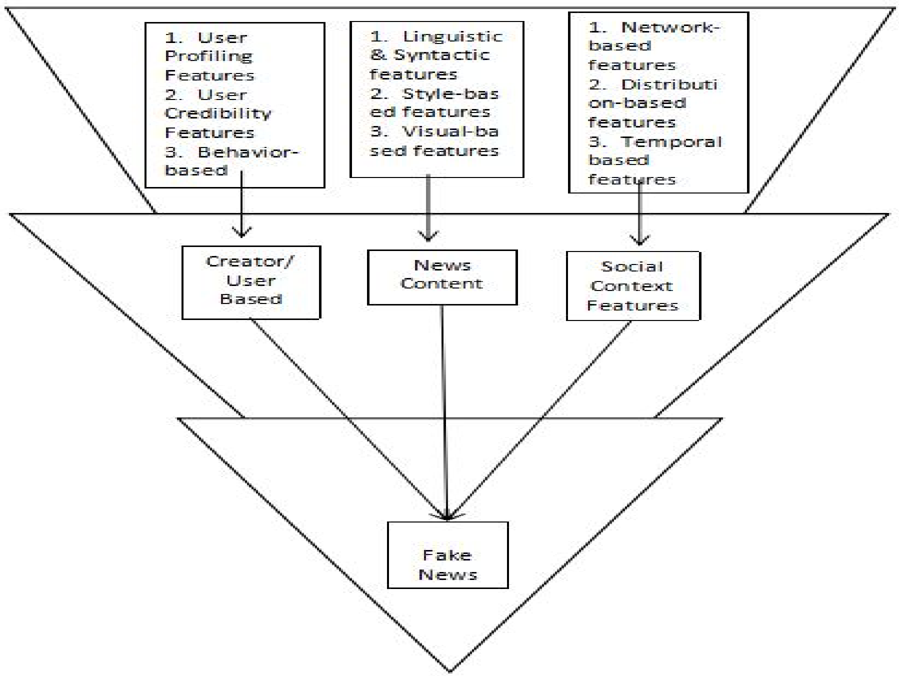

Fig 2. Conceptual Framework

21) Temporal-based features: Temporal-based features can be utilized to depict the posting conduct of online news creator in a period arrangement way.

They are great ascribing to distinguish suspicious posting exercises, and can be utilized to demonstrate the bogus degree of online news. Figure 2.2 below shows the various features of which fake news could be detected from which includes creator/user-based feature set, news content-based feature set, and social context-based feature set. And in each type of features exist different categories

D. Procedures for Developing Measurement Model for Fake News Conceptual Framework

In the development of the measurement model a questionnaire was distributed via Google form to various individuals to fill from various locations. The questionnaire was designed based on the conceptual model developed to know about individual's knowledge in respect to detecting fake news on various platform and self awareness. In this questionnaire a scale of 5 point was used to the level of different opinion giving by different individuals.

The format of the response is as follow:

Strongly Agree (SA)
Agree (A)
Undecided (U)
Disagree (D)
Strongly Disagree (SD)
Where each question are based on the conceptual model where we have three major features in respect to the fake news detection, which are
i. Creator/user based features
ii. News content based
iii. Social context based feature.
The creator/user based features (x) can be categorized into three which are
A. User profiling (x_1)
B. User credibility (x_2)
C. Behaviour based (x_3)
$X=f(x_1,x_2,x_3....x_n)$
The news content based (y) can be categorized into three which are
A. Linguistic syntactic features (y_1)
B. Style based feature (y_2)
C. Visual based features (y_3)
$Y=f(y_1,y_2,y_3........y_n)$
The social context based features (z) can be categorized into three which are
A. Network based (z_1)
B. Distribution based (z_2)
C. Temporal based (z_3)
$Z= f(z_1,z_2,z_3.......z_n)$

E. Differentiation between a Conceptual Framework And a Measurement Model

A conceptual framework which can also be referred to as a theoretical system is a scientific device with a few varieties and settings. It very well may be applied in various classes of work where a general picture is required. It is utilized to make applied differentiation s and sort out thoughts. Solid applied systems catch something genuine and do this in a way that is anything but difficult to recollect and apply. [5]Isaiah Berlin utilized the analogy of a "fox" and a "hedgehog" to make applied differentiation s in how significant thinkers and creators see the world. A measurement model also describe as structural equation model (SEM) is a type of causal displaying that incorporates a differing set of numerical models, PC calculations, and measurable techniques that fit systems of develops to information.

SEM incorporates corroborative factor examination, corroborative composite investigation, way investigation, halfway least squares way displaying, and idle development demonstrating. The idea ought not to be mistaken for the related idea of basic models in econometric, nor with auxiliary models in financial matters. Auxiliary condition models are frequently used to survey imperceptible 'inactive' develops. They frequently conjure a measurement model that characterizes inert factors utilizing at least one watched factors, and a basic model that ascribes connections between idle factors. [13]

III. RESULTS AND DISCUSSION

A. Result Analysis

The relationship between user profiling feature and creator/user features based on the question asked in respect to the questionnaire 52.9% agrees that user profiling feature can aid the spread of fake news. While 3.9% disagree, 39.2% strongly agreed while 3.9% where undecided.

Table I: Fake News Can Be Created/Spread by Individuals Who Has Large Number of Followers or Post on Social Media or Depending on Their Geo-Location

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	27	52.9	52.9	52.9
	Disagree	2	3.9	3.9	56.9
	Strongly Agree	20	39.2	39.2	96.1
	Undecided	2	3.9	3.9	
	Total	51	100.0	100.0	

The relationship between User Credibility Features and creator/user features based on the question asked in respect to the questionnaire 45.1% agrees that user Credibility can aid the spread of fake news. While 5.9% disagree, 31.4% strongly agreed while 17.6% where undecided.

Table II: Traditional News Brands or the Number of Friends And Followers of a User Aid the Spread of Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	23	45.1	45.1	45.1
	Disagree	3	5.9	5.9	51.0
	Strongly Agree	16	31.4	31.4	82.4
	Undecided	9	17.6	17.6	
	Total	51	100.0	100.0	

The relationship between user Behavior-based features and creator/user features based on the question asked in respect to the questionnaire 52.9% agrees that Behavior-based feature can aid the spread of fake news. While 9.8% disagree, 15.7% strongly agreed while 21.6% where undecided.

Table III: Do You Feel That Social Bot(Cyborg) Aids The Spread Of Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	27	52.9	52.9	52.9
	Disagree	5	9.8	9.8	62.7
	Strongly Agree	8	15.7	15.7	78.4
	Undecided	11	21.6	21.6	
	Total	51	100.0	100.0	

The relationship between user Linguistic & Syntactic features and Content Base Features based on the question asked in respect to the questionnaire 35.3% agrees that Linguistic & Syntactic features can aid the spread of fake news. While 15.7% disagree, 31.4% strongly agreed while 17.6% where undecided.

Table IV: The Headline of News Content or an Ad Can Be Considered When Detecting Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	18	35.3	35.3	35.3
	Disagree	8	15.7	15.7	51.0
	Strongly Agree	16	31.4	31.4	82.4
	Undecided	9	17.6	17.6	
	Total	51	100.0	100.0	

The relationship between user Style-based features and Content Base Features based on the question asked in respect to the questionnaire 13.7% agrees that Style-based features can aid the spread of fake news. While 37.3% disagree, 9.8% strongly agreed, 7.8% strongly disagree while 31.4% where undecided.

Table V: The Keystrokes Features, Font Style, External Link or Typing Style of News Content Can Be Considered in the Detection Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	7	13.7	13.7	13.7
	Disagree	19	37.3	37.3	51.0
	Strongly Agree	5	9.8	9.8	60.8
	Strongly Disagree	4	7.8	7.8	68.6
	Undecided	16	31.4	31.4	
	Total	51	100.0	100.0	

The relationship between user Visual-based features and Content Base Features based on the question asked in respect to the questionnaire 29.4% agrees that Visual-based features can aid the spread of fake news. While 31.4% disagree, 15.7% strongly agreed, 3.9% strongly disagree while 19.6% where undecided.

Table VI: The Number of Images or Videos Seen In a News Article Can Be Considered When Detecting Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	15	29.4	29.4	29.4
	Disagree	16	31.4	31.4	60.8
	Strongly Agree	8	15.7	15.7	76.5
	Strongly Disagree	2	3.9	3.9	80.4
	Undecided	10	19.6	19.6	
	Total	51	100.0	100.0	

The relationship between user Network-based features and Social Context Features based on the question asked in respect to the questionnaire 29.4% agrees that Network-based features can aid the spread of fake news. While 35.3% disagree, 13.7% strongly agreed, 3.9% strongly disagree while 17.6% where undecided.

Table VII: Can the Possibility of News Posted by Different Sources Be Considered as Fake News?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	15	29.4	29.4	29.4
	Disagree	18	35.3	35.3	64.7
	Strongly Agree	7	13.7	13.7	78.4
	Strongly Disagree	2	3.9	3.9	82.4
	Undecided	9	17.6	17.6	
	Total	51	100.0	100.0	

The relationship between user Distribution-based features and Social Context Features based on the question asked in respect to the questionnaire 43.1% agrees that Distribution-based features can aid the spread of fake news. While 27.5% disagree, 3.9% strongly agreed, 3.9% strongly disagree while 21.6% where undecided.

Table VIII: News That Are Likely Posted, Replied to or Responded Between Users From Social Media Platforms Can Be Considered to Contain Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	22	43.1	43.1	43.1
	Disagree	14	27.5	27.5	70.6
	Strongly Agree	2	3.9	3.9	74.5
	Strongly Disagree	2	3.9	3.9	78.4
	Undecided	11	21.6	21.6	
	Total	51	100.0	100.0	

The relationship between user Temporal-base and Social Context Features on the question asked in respect to the questionnaire 17.6% agrees that Temporal-base can aid the spread of fake news. While 31.4% disagree, 9.8% strongly agreed, 7.8% strongly disagree while 33.3% where undecided.

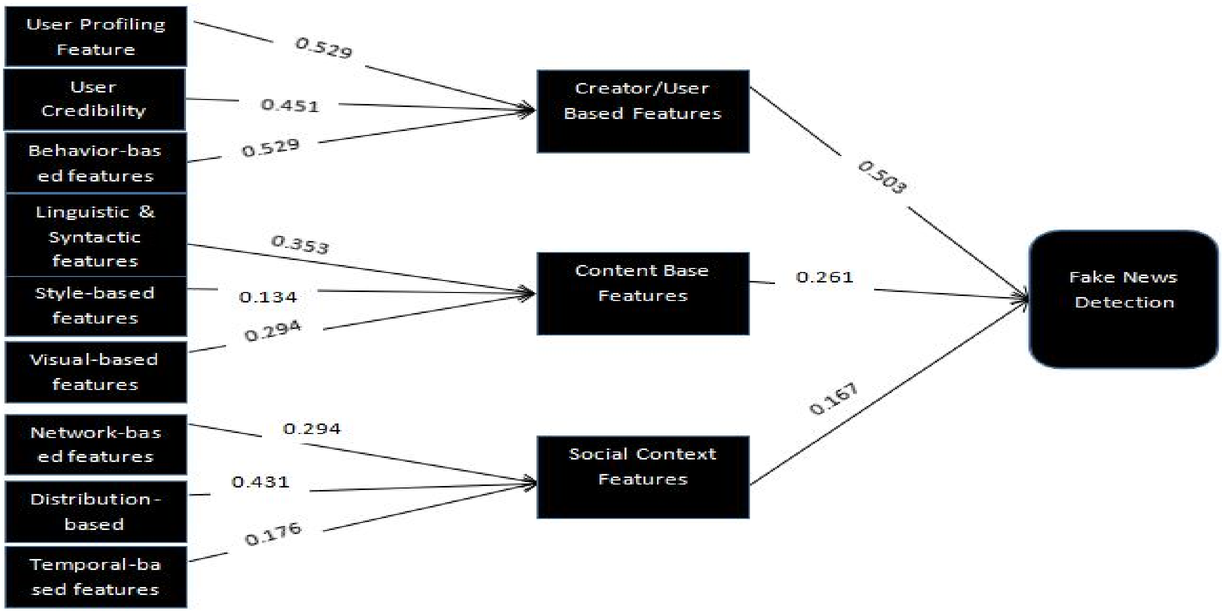

Fig 3. Diagram of Conceptual Framework with Weighted Links

Table IX: The Interval in Which Articles Are Been Posted, the Reply of Such Post From a Source Can Be Considered as Fake News.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	9	17.6	17.6	17.6
	Disagree	16	31.4	31.4	49.0
	Strongly Agree	5	9.8	9.8	58.8
	Strongly Disagree	4	7.8	7.8	66.7
	Undecided	17	33.3	33.3	
	Total	51	100.0	100.0	

Fig 3. shows a simple illustration in a diagram form the questions which were asked and analyzed in the table above, it shows the weighted links from the question based on the features set of fake news identification.

B. Measurement Model.

1. Creator/user based feature expressed as (X)

Given x_1, x_2, x_3 as User profiling, User credibility and Behaviour based respectively.

For: $x = f(x_1, x_2, x_3 ... x_n) : Ef(52.9, 45.1, 52.9)$

$f(x)/n$

$x = 50.2\%$

2. News Content based feature expressed as (Y)

Given y_1, y_2, y_3 as Linguistic syntactic features, Style based feature and Visual based feature.

For: $y = f(y_1, y_2, y_3 ... y_n) : Ef(29.4, 13.7, 35.3)$

$f(y)/n$

$x = 26.1\%$

3. Social Context Based feature expressed as (z)

Given z_1, z_2, z_3 as Network based, Distribution based, Temporal based.

For: $z = f(z_1, z_2, z_3 ... z_n) : Ef(45.1, 45.1, 45.1)$

$f(z)/n$

$x = 45.1\%$

The measurement model in fig 4 depicts the expression above

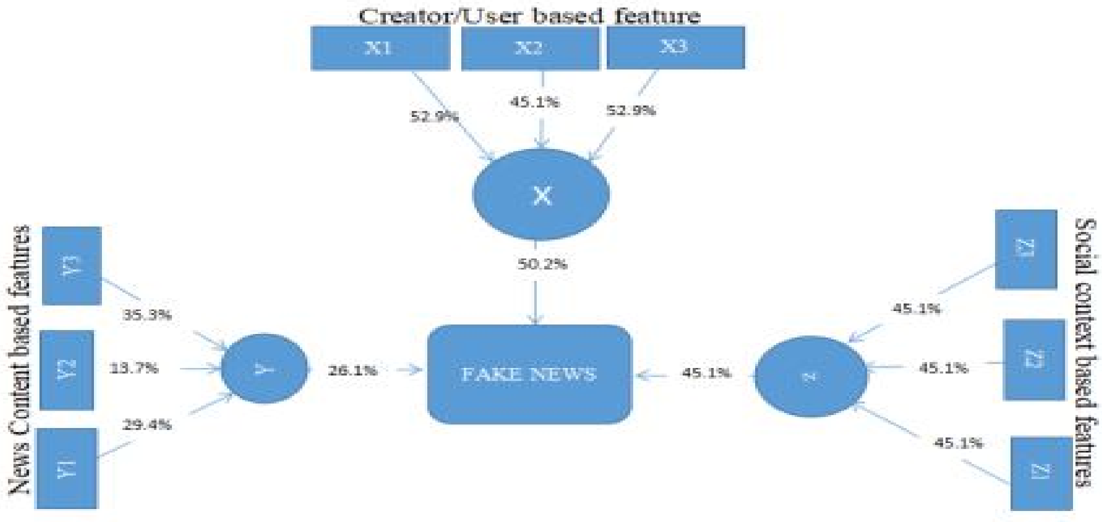

Fig 4: Measurement Model.

C. Performance Evaluation of Fake News Detection Models

Evaluate framework and measurement model: In evaluating the framework and measurement model, we are going to consider the framework based on the measurement model. Starting with **creator/user based features** it has three major categories: user profiling features, user credibility feature and behaviour based feature. Where the user profiling features have a total amount of 52.9% while user credibility is 45.1 and behavioural bases feature is 52.9. Having the highest percentage of 52.9%. user profiling and behavioural based is a way we could use to detect fake news faster in creator /user based.

Content based features contain three major categories: Linguistic and Syntactic features, Style based feature and Visual based feature. Where linguistic and syntactic features have a total amount of 35.3%, while style based feature is 13.4 and visual based feature is 29.4. Having the highest percentage of 35.3%. Linguistic and syntactic features are way we could use to detect fake news faster in content based.

Social context feature has three major categories: Network based feature, Distribution based features and Temporal based features. Where network based feature have a total amount of 29.4%, while distribution based features is 43.1and temporal based features is 17.6% Having the highest percentage of 43.1% distribution based feature is a way we could use to detect fake news faster. Hence to get which among the three features, creator/users based, content based feature and social context feature which could be used to detect fact news. In other to do this for creator/user based feature we got the average resulting to 50.3. Therefore creator/user based has a percentage of 50.3%. For Content bases average was 26.1 therefore content base have a percentage of 26.1%. For social context feature average was 16.7. Therefore social context have a percentage of 16.7%, According to the results it can be said that fake news can be easily spotted when it involves the creator/ user based feature as it has the highest percentage.

D. Comparative Analysis of Classifiers:

The data source used for this comparison is LIAR dataset. [15] The original dataset contained 13 variables/columns for train sets as follows: Column 1: the ID of the statement ([ID].json).

Column 2: the label. (Label class contains: True, Mostly-true, Half-true, Barely-true, FALSE, Pantsfire)
Column 3: the statement.
Column 4: the subject(s).
Column 5: the speaker.
Column 6: the speaker's job title.
Column 7: the state info.
Column 8: the party affiliation.
Column 9-13: the total credit history count, including the current statement.

9: barely true counts.
10: false counts.
11: half true counts.
12: mostly true counts.
13: pants on fire counts.

Column 14: the context (venue / location of the speech or statement)
Based on figure 2 in the measurement model which is based on the question in the questionnaire it is indicated that fake news is mostly spotted in the creator/user based feature, dues this we have chosen only 3 variables from this original dataset which indicates this aspect for the classification.
Below are the columns used in the dataset that have been used in this project.
Column 1: The speaker.
Column 2: The speaker's job title.
Column 3: The context (venue / location of the speech or statement)

Newly created dataset has only 2 classes as compared to 6 from original classes.
This columns selected where gotten based on the fake news questions used earlier in this project which indicated that fake news can be spotted mostly in the creator/user based feature. Analysis where taken based on feature selection and table shows the result obtained based on the different classifiers used on the dataset.

The result as presented in table 10 shows that **lazy ibk** performed better than others in terms of true positive rate (0.773%), false positive rate (0.240%), Accuracy (77.2786%), Precision (0.773%), Recall (0.0.773%), F-measure (0.772%), Area under ROC curve (0.817%). This was closely followed by Smo that achieved 0.740 as true positive rate, 0.284 as false positive rate, Accuracy of 73.9583%, Precision (0.742%), Recall (0.740%), F-measure (0.735%), Area under ROC curve (0.728%),

Table X: Comparative Analysis of Classifiers for Fake News Detection

Classifier With feature selections	TPR	FPR	Accuracy	Precision	Recall	F-Measure	Roc Area
SMO	0.740	0.284	73.9583	0.742	0.740	0.735	0.728
Lazy Ibk	0.773	0.240	77.2786	0.773	0.773	0.772	0.817
Naïve Bayes	0.712	0.304	71.224	0.711	0.712	0.710	0.774
Decision Table	0.642	0.370	64.1927	0.641	0.642	0.641	0.695
Svm	0.601	0.446	60.1237	0.594	0.601	0.586	0.578

Based on the dataset used in this project it shows that when feature selection is used in the classification we have better accuracy based on identifying fake news from the creator/user based feature compare to when all the features were used. This shows the importance of feature selection for the enhancement of the performance of the classifier. Using feature selection we achieved the following performance based on Lazy Ibk: True Positive Rate (0.773%), false positive rate (0.240%), Accuracy (77.2786%), Precision (0.773%), Recall (0.0.773%), F-measure (0.772%), Area under ROC curve (0.817%).

Table XI: Feature Selection Comparison against Non Feature Selection

Classifier	**TPR**	**FPR**	**Accuracy**	**Precision**	**Recall**	**F-Measure**	**Roc Area**
Random Forest Without feature Selection	-	-	61.47	-	-	-	-
Lazy Ibk With Feature Selection	0.773	0.240	77.2786	0.773	0.773	0.772	0.817

Using the creator/user based feature as a feature selection it outperformed the Random forest classifier without feature selection in term of Accuracy.

IV. CONCLUSIONS

Recently, fake news is becoming one of the most threatening harms on social media. Fake news can be used by malicious entities to manipulate people's options and decisions on important daily activities, like stock markets, health-care options, online shopping, education, and even presidential election. Although there are incentives for posting and spreading fake news, there seems almost no incentive to overcome the cost for preventing and checking fake news. Society needs time to reach equilibrium in the trade-off between freedom to post and public wealth lost by fake news. In this paper we discuss about a conceptual framework that could used in detecting fake news where we said it can be as Creator/user based features, News content based, Social context based feature

ACKNOWLEDGMENT

I thank God Almighty for His grace over me during this research, may his name be praised. Giving a special appreciation to my supervisor specially Dr Joseph.A. Ojeniyi for what he has impacted on me during this work. I also want to thank Parents for all their immeasurable support and encouragement given to me during the research, May God continue to bless you all in Jesus name, AMEN. Special thanks the lecturers of cyber security who, out of their busy schedules, took time to advise me in my drafting. I am grateful to all my colleagues who contributed to my research during this period, thank you and God bless you.

REFERENCES

1. Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. Procedia Computer Science, 141, 215–222. https://doi.org/10.1016/j.procs.2018.10.171
2. Al-Ash, H. S., & Wibowo, W. C. (2018). Fake news identification characteristics using named entity recognition and phrase detection. Proceedings of 2018 10th International Conference on Information Technology and Electrical Engineering: Smart Technology for Better Society, ICITEE 2018, 12–17. https://doi.org/10.1109/ICITEED.2018.8534898
3. Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying Fake News and Fake Users on Twitter. Procedia Computer Science, 126, 451–461. https://doi.org/10.1016/j.procS.2018.07.279
4. Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). News in an Online World : The Need for an " Automatic Crap Detector " News in an Online World : The Need for an " Automatic Crap Detector ," (September).
5. "Conceptual framework", Wikipedia, 2019-10-24, retrieved 2019-10-31
6. Figueira, Á., & Oliveira, L. (2017). ScienceDirect ScienceDirect The current state of fake news : challenges and opportunities The current state of fake news : challenges and opportunities. Procedia Computer Science, 121, 817–825. https://doi.org/10.1016/j.procs.2017.11.106
7. Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. 2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings, 900–903. https://doi.org/10.1109/UKRCON.2017.8100379
8. Jain, A., & Kasbe, A. (2018). Fake News Detection. 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2018, 1–5. https://doi.org/10.1109/SCEECS.2018.8546944
9. Kotteti, C. M. M., Dong, X., Li, N., & Qian, L. (2018). Fake news detection enhancement with data imputation. Proceedings - IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, IEEE 16th International Conference on Pervasive Intelligence and Computing, IEEE 4th International Conference on Big Data Intelligence and Computing and IEEE 3, 193–199. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00042
10. Knshnan, S., & Chen, M. (2018). Identifying tweets with fake news. Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018, 67, 460–464. https://doi.org/10.1109/IRI.2018.00073
11. Shabani, S., & Sokhn, M. (2018). Hybrid machine-crowd approach for fake news detection. Proceedings - 4th IEEE International Conference on Collaboration and Internet Computing, CIC 2018, 299–306. https://doi.org/10.1109/CIC.2018.00048
12. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600
13. "Structural equation modeling", Wikipedia, 2019-10-25, retrieved 2019-10-31
14. Vedova, M. L. Della, Tacchini, E., Moret, S., Ballarin, G., & Dipierro, M. (2018). Automatic Online Fake News Detection Combining Content and Social Signals. 2018 22nd Conference of Open Innovations Association (FRUCT), 272–279.
15. William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, to appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), short paper, Vancouver, BC, Canada, July 30-August 4, ACL.
16. Zhang, X., & Ghorbani, A. A. (2019). An overview of online fake news : Characterization , detection , and discussion. Information Processing and Management, (August 2018), 1–26. https://doi.org/10.1016/j.ipm.2019.03.004