FAKE NEWS DETECTION USING LINGUISTIC INQUIRY WITH SYLLABLE PATTERN OF WORDS BASED ON CNN APPROACH

Zahir Abbas Khan[1*], Rekha V[2], Ajit Danti[2], Naveen J[2]

Abstract:
The usage of social media for news consumption is progressively growing due to its accessibility, low cost, increased attractiveness, and capacity to disseminate "false news."Fake information deliberately created is purposefully or unintentionally created over the internet. Some people spread wrong information on social media to get the attention or financial and political gain. This is affecting a larger group of society who are blind by technology. We need to be smarter at the recognition of fake or real news. In this research we have implemented Linguistic Inquiry with Syllable Pattern of Word (LISPW) feature set for detection of fake news based onConvolutional neural network (CNN). The Linguistic features involved certain textual characteristics converted into a numerical form such that they can be used as an input for the training models. Tensorflow is used for implementation of the proposed framework and provides visualizations for the CNN. Confusion matrix and classification reports are validated and achieved accuracy score of 98% by using Linguistic Inquiry with Syllable Pattern of Word. Also, to check the headline or fact of the news, a module is developed to check for contradiction between inputted news and web scraped news related to it.

Keywords: Fake news, Feature Extraction, Deep Learning, Convolutional Neural Network, LISPW.

[1*]Research Scholar, Department of Computer Science and Engineering, School of Engineering and Technology, Christ (Deemed to be University), Bangalore, India
[2]Department of Computer Science and Engineering, School of Engineering and Technology, Christ (Deemed to be University), Bangalore, India

***Corresponding Author:** Zahir Abbas Khan
*Research Scholar, Department of Computer Science and Engineering, School of Engineering and Technology, Christ (Deemed to be University), Bangalore, India
Email: zahir.khan@res.christuniversity.in

DOI: 10.48047/ecb/2023.12.si10.0024

1. INTRODUCTION

Social media is a platform that enables the sharing of user-generated content, information, ideas, and expression in the quickest, easiest-to-access, and cost-free manner possible. Kai Li and others (2020). It has evolved into one of the simplest, earliest, and best sources of news in the modern era. Abhinandan Chakraborty and others (2019).

Due to the lack of an administration body on social media, it frequently spreads fake news and allows for the distribution of material that is of lesser quality than the actual news. Compared to traditional news sources, news with photos and videos grabs the reader's attention and develops a better storyline. Although the internet and technology have collectively made life easier in many ways, social media has also given rise to false information. Fake news is the dissemination of false information made possible by inborn human inclinations.

Fake news denotes a type of yellow press which intentionally presents misinformation or hoaxes spreading through both traditional print news media and recent online social media Jiawei Zhang et. al. (2019). Fake news has been existing for a long time, since the "Great moon hoax" published in 1835. Counterfeit news or fake news is a bit of false data created for business enthusiasm to pick up consideration and produce promotion income or to spread scorn related violations to impact the world politically Kyeong-Hwan Kim et. al.(2019). Supposed accurate news items that purposefully misrepresent reality in order to pique attention, increase viewing, or otherwise deceive. There have been many instances recently of unverified or fraudulent information spreading swiftly through online informal organizations.

According to a post-election statistical report, online social networks account for more than 41.8% of the fake news data traffic in the election, which is much greater than the data traffic shares of both traditional TV/radio/print medium and online search engines respectively Madhu Nakerekanti et. al. (2019). An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely.

Everyday access of news sources, for example, web based life channels, news sites, and online papers have created demanding to verifying reliable news sources due to enhancement of misleading information. We center around the ID of phony contents or articles in news sites. First, we present database for the phony news discovery task, using numerous news spaces and depict the accumulation, explanation, and approval process in detail and present a few exploratory examinations on the acknowledgment of etymological varieties in phony and real news content.

Detecting fake news from social media is a quirky challengeKai Shu et. al. (2018). In addition to information about news content; user engagement and user's social behavior on social media needs to be explored. It is obvious to have a comprehensive dataset including news content, social context and spatio-temporal information to ease the research CagdasBak et. al. (2018). It is a challenge to have a dataset with spatio-temporal information which signifies how fake news generates over time, how the user reacts and how to extract useful temporal patterns for early interventions.

The Overall Objective of the project 'Fake news detection' is to classify the news article or other documents into certain or not. Classification is the method to forecast the label which is unidentified before to distinguish between one object to another on the basis of selected feature or attributes Wee Jing Tee et. al. (2018). In this method, data will be divided into two part first one is training data i.e. information to be related to find out the category label, second one testing data where we perform the test to know the category label of the new object. We explore identification of fake news using various models and classifiers and predict the accuracy of different models and classifiers. Through this project, we examine which model will give more accuracy and classify the news into real or fake.

The remaining paper is organized as follows: Section II relates the literature survey, machine learning techniques for fake news detection method is introduced in Section III. Experimental results are discussed in Section IV and finally Section V concludes the paper.

2. LITERATURE SURVEY

Through this section, we summarize some of the existing research works in the field of Machine learning/deep learning to analyses about Fake News Detection and build a model according to the existing applications.

H. Ahmed et al. (2017) extracted linguistic features such as n-grams from textual articles and training multiple ML models including K-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD), achieving the highest accuracy (92%) with SVM and logistic regression. According to the research, as the number of n increased in n-grams calculated for a particular article, the overall accuracy decreased). Thephenomenon has been observed for learning models that are used for classification.

K. Shu et al. (2017) achieved better accuracies with different models by combining textual features with auxiliary information such as user social engagements on social media. The authors also discussed the social and psychological theories and how they can be used to detect false information online. Further, the authors discussed different data mining algorithms for model constructions and techniques shared for features extraction. These models are based on knowledge such as writing style, and social context such as stance and propagation.

M. Granik et. al.(2017)presents a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. We achieved classification accuracy of approximately 74% on the test set which is a decent result considering the relative simplicity of the model. These results may be improved in several ways that are described in the article as well. Received results suggest, that fake news detection problem can be addressed with artificial intelligence methods.

Conroy et. al. (2015)mainly focuses on categorizing the news based on finding the degree of accuracy or correctness in the news. Include mainly two categories for assessmentlinguistic cue approach (with machine learning) and network analysis approach. Both approaches adopt machine learning techniques for training classifiers to suit the analysis. These papers show the current development of correctness assessment methods, their goals and classes with the aim to propose new hybrid system for detection.

Pasquini et. al. (2015) introduces a paper which is huge impact on the deception of users, whose opinion can be seriously influenced by altered media. In this work, we face the challenge of verifying online news by analyzing the images related to the particular news article. Our goal is to create an empirical system which helps in verifying the consistency of visually and semantically similar images used within different news articles on the same topic. Given a certain news online, our system identifies a set of images connected to the same topic and presenting common visual elements, which can be successively compared with the original ones and analyzed in order to discover possible inconsistencies also by means of multimedia forensics tools.

Gupta et al. (2012) have assessed the credibility of twitter experiences with Page-Rank approach for credibility analysis. Identification of classifier approach, the credibility analysis and event graph-based optimization approach presented. Castilloet al. (2011) have investigated with four types of features: message-based features, user-based features, topic-based features and propagationbased features to classify a tweet to be credible or not

3 FAKE NEWS DETECTION METHOD

The block diagram of Fake news detection method using Linguistic Inquiry with Syllable Pattern of Word (LISPW) based on Convolutional neural network (CNN) is represented in below Fig. 1. The first step in this classification problem is dataset collection phase, followed by preprocessing, implementing feature extraction, and then performs the training and testing of dataset and finally running the classifiers.

FakeNewsNet is a multi-dimensional data repository that currently contains two datasets with news content, social context, and spatiotemporal information. The dataset is constructed using an end-to-end system, FakeNewsTracker. We describe and compare FakeNewsNet with other existing datasets in FakeNewsNet: A Data Repository with News Content, Social Context and Spatial temporal Information for Studying Fake News on social media. The proposed method has been evaluated on Gossipcop and PolitiFact dataset

from the FakeNewsNet Dataset. The dataset contains the following information:

- Real and fake news content: Contains news articles with attributes such as news id, title, text, URL, authors and source.
- News and user engagement: Specifies the number of times a news article has been shared by a user.
- User–user relationship: Specifies the user network on social media.

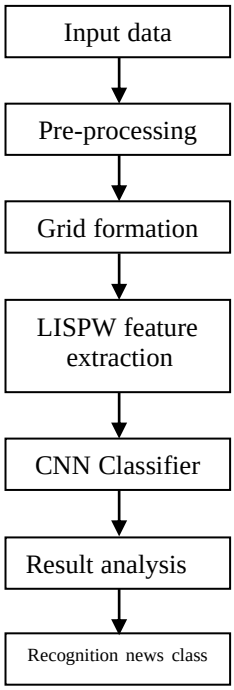

Fig.1:Processblockdiagram

After the acquisition of content, pre-processing is carried out. Which includes the following steps:

- All letters in the document are converted to lowercase
- Numbers are removed
- Punctuations, accent marks are removed
- White spaces are removed
- Stop words are expelled.

The LISPW method extracts the features of given data by using the syllable pattern with various projection points. This separates the content in the statement with relevant attributes to form the better learning database compare to other traditional data retrieval methods. To overcome the problem and to enhance the classification performance, a novel feature extraction method with the Deep Learning (DL)based classification model can be implemented to retrieve best attribute among the overall database.CNN along with Linguistic Inquiry with Syllable Pattern of Word (LISPW) feature set used in this research are the novelty of our proposed approach. Linguistic features involved certain textual characteristics converted into a numerical form such that they can be used as an input for the training models. These features include percentage of words implying positive or negative emotions; percentage of stop words; punctuation; function words; informal language; and percentage of certain grammar used in sentences such as adjectives, preposition, and verbs.

At initializing phase,

$$y(n) = \{y_0, y_0 \ldots . y_0\} \ \forall \ n = 1,2,3, \ldots . m \ldots (1)$$

Where, 'm' represents the number of features.

$$\omega_i(n) = \left(\frac{q_i^n(y(n))}{\sum_{i=1}^{N}(\, q_i^n(y(n)))}\right) \forall \ n = 1,2,3, \ldots . m \ldots \qquad (2)$$

Where, $\omega_i(n)$ represents the weight value of attributes for i^{th} iteration.

$$L_{1:i}^m = L_{1:i-1}^m \times L_i^m \ldots . . \qquad (3)$$

Where, $L_i^m = \frac{1}{N}\sum_{i=1}^{N} P_i^n(y(n))$,

$$P_i^n = e^{\{-\sum y \in T_D f_i(y)\}}$$

Weight value update as,

$$\omega_i(n+1) = \sum_{l=1}^{N} \omega_i(n)\, \delta(x_n) \ldots . \qquad (4)$$

Update Attributes,

$$y(n+1) = \frac{1}{N}\sum_{l=1}^{N} \delta(y_n) \ldots\ldots \quad (5)$$

$$m_i^* = max(L_{1:i}^m) \ldots\ldots \quad (6)$$

$$\omega_i^*(n) = max\left(P_i^n\big(y(n)\big)\omega_i(n)\right) \ldots\ldots (7)$$

Where, the algorithm of LISPW is as follows below:

Algorithm 1: Linguistic Inquiry with Syllable Pattern of Word (LISPW)Algorithm

1.Input: Input DataT_D, Output: Features of attributes $F_D(s)$.For$i$ = 1 to M //Loop run for 'M' number of iteration.
2. Initialize attributes 'y'and the weight value 'α'
3. Estimate the likelihood of the attributes by $L_{1:i}^m$.
4. Update weight value and Update Attributes.
5. Find maximum likelihood, m_i^*.
6. Find maximum relevance value, $\omega_i^*(n)$.
7. Update weight value of attributes and get best relevance value to form feature set.
If $(L_{1:i}^m) > 0$, then
$s_v = \{s_{v-1},\ i\ \}$

End if
8. ElseContinue for loop 'i'.
End If
$F_D(s) = T_D(s_v)$
End 'i' Loop

Deep learning classifier operation includes the following expressions as,

$$F_D(s) = \{T_{D1}(s),\ T_{D2}(s),\ \ldots,\ T_{Dm}(s)\} \ldots \quad (8)$$

In the input layer of classifier, the data sequence can be formed as the matrix as in below equation.

$$X_D(s) = \begin{bmatrix} F_{D1}(s) \\ F_{D2}(s) \\ \cdots \\ F_{Dm}(s) \end{bmatrix} \ldots. \quad (9)$$

Form the matrix arrangement, the block correlation feature can be estimate by,

$$F\big(X_D(s).X_D^*(s)\big) = X_D^*.e^{T-T_m} \ldots \quad (10)$$

Where, 'T' and 'T_m' represents the attribute values from matrix $X_D(s)$.

$$K_m = \frac{1}{2^{q-1}}\left(\frac{\sqrt{2q}}{l}\right)^q k_q\left(\frac{\sqrt{2q}}{l}r\right)\forall q = 1,\ 2,\ \ldots,\ N \ldots (11)$$

Where, r' represents range of featuredistance, 'l' represents the length of feature vector.
Texture relevancy as
$t_n = F^T\omega_n$; 'ω_n' weight value of attributes.

Extract the training features and form the network by

$$X_b = \overline{X_b} + \sum_{i=1}^{N} t_i(d)p^i \ldots \quad (12)$$

$$\widehat{T_s} = \left(\left(X_b^{\bar{d}} - \overline{X_b}\right)^T(P^T)\right)^T \ldots. (13)$$

Where, the relevance factor $X_b^{\bar{d}} \in R^{(T-T_p)M}$ can be written as

$$R^{(T-T_p)M} = \hat{T_s}^T Q^T + \overline{tt_a} \ldots\ldots \quad (14)$$

Where, 'P' and 'Q^T' – Predicted component. The predicted label can be representing by

$$V(k) = \frac{d_{ij}}{R_j - R_i} \ldots \quad (15)$$

Where, d_{ij} – Distance matrix for 'i' and 'j' of the relevance matrix 'R'

The algorithm of Deep Learning(CNN Classifiers)is as follows below:

Algorithm 2: CNN classifier

1. Input: Training set $F_D(s)$
2. Output: Classified Result$V(k)$
3. Arrange the input series in the sequential order $F_D(s)$.
4. Estimation of block correlation feature $F\left(X_D(s).X_D^*(s)\right)$
5. Estimate the kernel model of classifierK_m.
6. Estimate the relevancy using kernel function with feature points.
7. Extract the training features and form the network.
8. Estimate the matching score for the correlated blocks by $\hat{T_s}$.
9. End

This work implements the aforementioned a different model in Python and evaluates its performance on two real-world fake news datasets. The following subsections describe the datasets in detail; explain the implementation decisions and the comparison baselines, as well as the achieved results. The baseline methods comprise state-of-the-art techniques for fake news classification and a few more newly produced base lines.

4. EXPERIMENTAL RESULTS

The proposed method has been evaluated on Gossipcop and PolitiFact dataset from the FakeNewsNet Dataset. Next, each dataset is spilt into training and test subsets (80–20% split). For the evaluation of results, four metrics have been used, which are based on the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) in the predictions of the binary classifiers:

Accuracy:The capacity of the framework to precisely characterize information depends to a vast degree on the illustrations that you give.

5. $Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \ldots (16)$

Precision also called positive predicted value is the fraction of significant instances among the retrieved instances.

$$Precision = \frac{TP}{TP + FP} \ldots \quad (17)$$

Recall also known as sensitivity or True Positive Rate (TPR) is the fraction of significant instances that have been retrieved over the total amount of relevant instances.

6. $Recall = \frac{TP}{TP+FN} \ldots$ (18)

F1 score also F-score or F-measure is a measure of a test's accuracy for binary classification.

7. $F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \ldots (19)$
8. True Negative Rate or Specificity is defined as follows:
9. $\text{Specificity} = \frac{TP}{TP+FN} \ldots$ (20)

Table 1: PERFORMANCE OF DIFFERENT CLASSIFIERS

Parameters	LSTM (%)	KNN (%)	LISPW Algorithm based detection (%)
Accuracy	96.75	95.2	98.21
Precision	98.01	96.4	98.99
Recall	95.41	94.4	100
F1-score	97.26	96.5	99.93
Specificity	99.05	92.2	99.74
Kappa coefficient	96.1	94.8	97.8

Fake News detection using different classifiers as LSTM, K-nearest Neighbor (KNN) and described LISPW performance is elaborated in Table 1. From Table 1, it is found that performance parameters are comparatively high in LISPW Algorithm based fake news detection method than the remaining two classifiers. Graphical representation of performance of LSTM and LISPW based fake detection is represented in below Fig. 2.

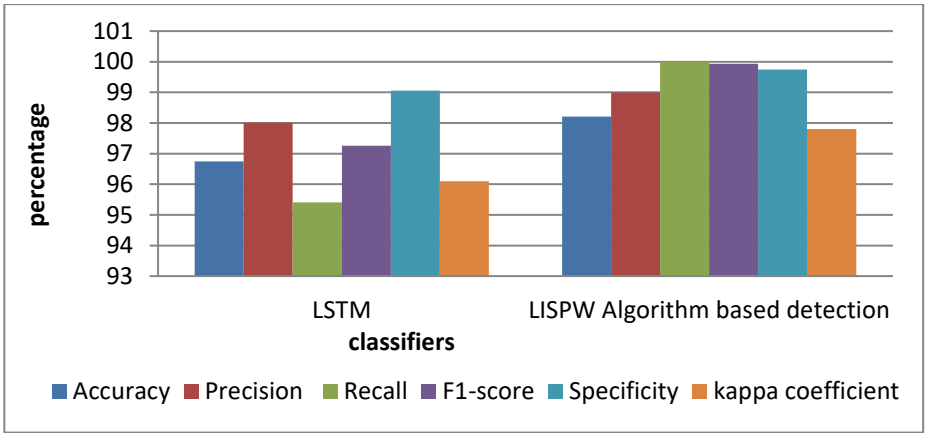

Fig. 2: PERFORMANCE COMPARISON FOR METHODS USING DIFFERENT PARAMETERS

The relation between False Positive Rate (FPR) and True Positive Rate (TPR) is given in below Table 2 and its graphical representation as shown below Fig. 3.

FPR	TPR		
	LSTM	KNN	LISPW Algorithm based detection
0	0	0	0
0.1	0.7	0.7	0.5
0.2	0.8	0.8	0.6
0.85	1	0.9	0.7
0.95	1	1	1
1	1	1	1

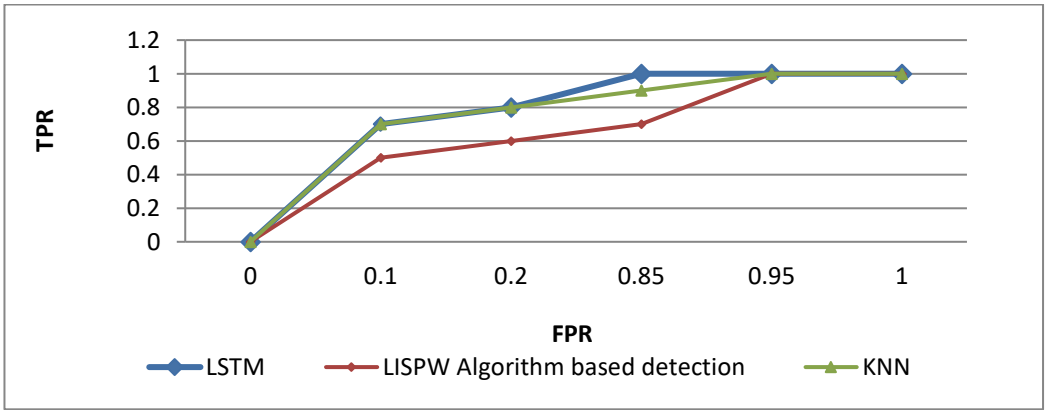

Fig. 3: FPR VERSUS TPR GRAPH

Error rate, kappa coefficient and FPR of LSTM and LISPW Algorithm based detection methods are described in below Table 3 and graphical representation is shown in Fig. 4.

Table 3: FPR AND ERROR RATE COMPARISON

Parameters	LSTM	LISPW Algorithm based detection
FPR	0.012	0.015
Error rate	0.04	0.02

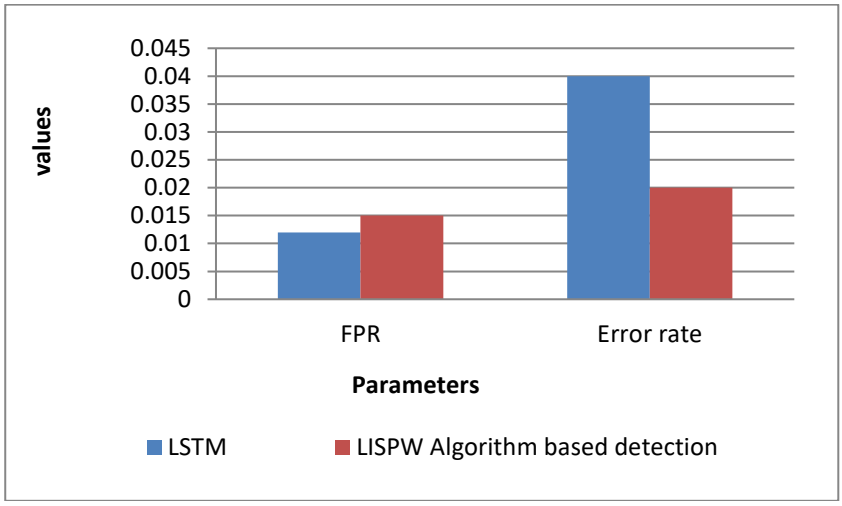

Fig. 4: FPR AND ERROR RATE VALUES FOR DIFFERENT METHODS

These metrics are commonly used in the machine learning community and enable us to evaluate the performance of a classifier from different perspectives. Specifically, accuracy measures the similarity between predicted fake news and real fake news.From above results it is clear that, described LISPW Algorithm based fake news detection method is works efficiently than the LSTM method.

5. CONCLUSION

Fake News Detection is the analysis of socially relevant data to distinguish whether it is real or fake. In this work, Linguistic Inquiry with Syllable Pattern of Word (LISPW) feature set for detection of fake news based on Convolutional neural network (CNN) method is described.In this work, the aim to implement feature extraction technique for better Representation of the data and investigate the authenticity of the news on social media as legitimate or fake news and provide a solution to detect fake news. First, we will investigate the potential and foundation of other types of different dataset of real and fake news in a similar way, such as content features and social network features, for fake news detection. Second, we will further investigate the correlations between data cleaning and data analysis for the fake news pieces. Then various parameters are considered for performance evaluation and these are Accuracy, Precision, Recall, F1-Score, Specificity, Kappa coefficient, error rate and False Positive Rate (FPR). From results it is clear that, described LISPW Algorithm based fake news detection method is works efficiently than the LSTM and KNN classifications.

References

1. Abhijnan Chakraborty, Saptarshi Ghosh, NiloyGanguly, Krishna P. Gummadi, “Editorial Versus Audience Gatekeeping: Analyzing News Selection and Consumption Dynamics in Online News Media”,IEEE Transactions on Computational Social Systems, Volume: 6, Issue: 4, Year: 2019.
2. CagdasBak, AysunKocak, ErkutErdem, AykutErdem, “Spatio-Temporal Saliency Networks for Dynamic Saliency Prediction”, IEEE Transactions on Multimedia, Volume: 20, Year: 2018
3. C. Castillo, M. Mendoza, and B. Poblete, “Proceedings of the 20th international conference companion on world wide web, www 2011,” in conference, 2011.
4. Conroy, N. J., Rubin, V. L., & Chen, Y. “Automatic deception detection: Methods for finding fake news”, Proceedings of the

Association for Information Science and Technology, 52(1), 1-4, 2015
5. H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," Security and Privacy, vol. 1, no. 1, 2017.
6. Jiawei Zhang, Bowen Dong, Philip S. Yu, "Deep Diffusive Neural Network based Fake News Detection from Heterogeneous Social Networks", 2019 IEEE International Conference on Big Data (Big Data), Year: 2019
7. Kai Shu, Suhang Wang, Huan Liu, "Understanding User Profiles on Social Media for Fake News Detection", 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Year: 2018
8. Kai Li, Lin Huang, Zeyuan Song, "Understanding the Emergence of Collaborative Problem-Solving Practices in Enterprise Social Media: The Roles of Social Factors", IEEE Access, Volume: 8, Year: 2020
9. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
10. Kyeong-Hwan Kim, Chang-Sung Jeong, "Fake News Detection System using Article Abstraction", 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Year: 2019
11. Madhu Nakerekanti, V.B. Narasimha, "Analysis on Malware Issues in Online Social Networking Sites (SNS)", 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Year: 2019
12. M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), May 2017, pp. 900–903, doi: 10.1109/UKRCON.2017.8100379.
13. M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on twitter," in Proceedings of the 2012 SIAM International Conference on Data Mining, SIAM, 2012.
14. Pasquini, C., Brunetta, C., Vinci, A. F., Conotter, V., &Boato, G. "Towards the verification of image integrity in online news", In Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on (pp. 1-6), IEEE, 2015
15. Wee Jing Tee, Raja Kumar Murugesan, "Trust Network, Blockchain and Evolution in Social Media to Build Trust and Prevent Fake News", 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Year: 2018.
16. A. Gupta and P. Kumaraguru. Twitter explodes with activity in mumbai blasts! a lifeline or an unmonitored daemon in the lurking? IIIT, Delhi, Technical report, IIITD-TR-2011-005, 2011.
17. Shrutika S Jadav,Sudeep D Thepade"Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier", Applied Artificial Intelligence, An International Journal,Sep 2019, Taylor &Francis,ISSN: 0883-9514 (Print) 1087-6545 (Online).
18. Prof.Sunil Yadav, Siddharth Singh, Shripad Shukla, "Fake News Detection Using Machine Learning", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 10, Issue 1, October 2020, ISSN (Online) 2581-9429, ISSN (Print) 2581-942 X
19. Harita Reddy, Namratha Raj, Manali Gala AnnappaBasava, "Text-mining-based Fake News Detection Using Ensemble Methods" International Journal of Automation and Computing, December 2019, Springer-Verlag GmbH Germany, DOI: 10.1007/s11633-019-1216-5,