

# FAKE NEWS DETECTION TECHNIQUES

**Shashank Sahu**

*Professor, Ajay Kumar Garg Engineering College Ghaziabad*

**Abstract**—Fake news is increasing daily related to any topic in the social media. It creates a panic situation for society and gives benefits to few persons. In this digital world, creating and spreading the fake news is very easy and it is also affecting the people in large. Fake news is an information which is not truly right. It has been designed and created based on non-factual data to give benefit to few people. Fake news is disseminated using various media. It is convenient and fast to disseminate the fake news using digital media such as social media. Meantime, it also increases challenges to identify fake news. This paper discusses approaches used in detecting fake news and addresses various challenges in detecting the fake news.

## I. INTRODUCTION

The use of social media is increasing rapidly and hence spreading of fake news in this media is also increasing in fast pace [1]. Social media is used to develop relationship among people. It needs sharing of information. There are many ways from which users of social media may share their views and information. In which one of the methods is status checking. It does not require much skill for sharing the information [2]. Social media is open for everyone. It has users of both skilled and less skilled persons. Because of less skilled persons, it opens dissemination of information which is not true easily to be propagated in the social media [3]. However, skilled persons are also not away from being affected from the fake news.

The impact of fake has effects on popularity of parties. Fake news works in both ways: it may increase the popularity of the party, which is actually not true or it may defame the popularity of the party which is basically a false information about the party, which is away from the truth. In the last few years, electronic media is mostly used for disseminating the fake news specifically for political gain. The public opinion is polarised using propaganda of fake news [4].

Fake news and misinformation are promoting violent and hate speech. It undermines democracies and reduces trust in the democratic processes. Propaganda, misinformation and fake news are often overlap in the meaning. It is also referring to range of ways in which sharing information causes harm intentionally or unintentionally in relation to the promotion of moral and political causes.

## II LITERATURE SURVEY

Social media includes both websites which are developed for forum, website, microblogging and social bookmarking.

There is fake news that spreads widely related to health that exposes the health of global concern. Natural Language Processing is used to detect actions for various languages such as Italian, English, and Dutch speeches to recognize different languages. For example, detection and emotion analyser, and sentiment analysis are used to extract a particular topic for the subject [5][6]. Most research related to speech techniques generally performed for European language. Supervised method uses the training data to see patterns which are hidden. In case of unsupervised method, hidden pattern is recognized hidden data with the help of unlabelled data.

Machine learning [4] is a method that find out accurate result without reprogram the machine. Machine analyse the data and predict the output. When the training of the machine learning is completed, the algorithm starts work on the new data. Decision Tree is working best on flow chart. It is used for problems related to classification. Internal node is also a decision tree. The leaf node states label of the class which is achieved from moving from root to leaf node through internal node i.e. decision node.

Random forests are build using many decision tree algorithms [4]. Each decision tree gives separate result. Random Forest are constructed using various decision tree algorithms that gives separate result for each decision trees. The result which has been predicted by many decision trees is taken for the random forest. For providing variations in decision trees, the random forest is able to select randomly subcategory of features from each group. Support Vector Machine (SVM) algorithm has the space of n dimensional for plotting the data with help of coordinates which represents the feature value. In this case hyper plane is identified to make the two classes separate for data classification. Naive Bayes states that one function in the category does not depend on another Such as a fruit can be classified as an apple with properties it is swirls and red color, and diameter is near to 3 inches.

KNN classifies new position from neighbouring k with respect to them. It is a kind of supervised learning. It is mainly used for intrusion detection and for pattern recognition. In this case, no specific distribution is assigned to data.

In the last few years, electronic media has been used to disseminate the online political fake news. It is used to fabricate misinformation with politically charged content. Misinforma-

tion is also considered as to believe to support the electronic growth of far-right party. Fake checker is checking the fake news within three months from the preceding of the election. The result of voting has two implications firstly, fake news shows positive on the voting for populist candidates and that secondly, the fake news cannot be impacted most of the political party growth. Facebook page that observed increase in fake news while the support of political party [6] [7].

Fake news is threat in this information-centric age. Most of the studies focus on a particular domain and using limited algorithms. Performance of algorithms drops while detecting the fake news in various domain. The ensemble techniques perform better than individual technique.

Fake news consists of misleading of information [8]. It disseminates lies about certain statistics or services of the country. It is dealing some time by making author accountable for disseminating the information. Fig. 1 describes the methodology of detecting the fake news.

However, this is limited they depends on human manual interaction, it is difficult to detect millions of articles published and removed those articles which are fake on other's author name [9]. One of the solutions may the development of scoring of automated index or rating to different publishers, and news context. The other techniques to detect a news is fake or not only on basis of word, phrases, sentences and titles when applying techniques of supervised learning [10][11].

### III. APPROACHES FOR FAKE NEWS DETECTION

Following is the list of approaches in detecting the fake news

1. Approach of language
2. Approach of Topic-agnostic
3. Approach of Machine learning,
4. Approach of Knowledge-based,
5. Approach of Hybrid

**1. Approach of Language:** It is based on use of linguistic by software program or a human being in detecting the news of fake. In general, people who spreading the fake news have knowledge about how to write the fake news stories. But style of their language is exposed to outside. It uses all the words of a sentence. It also considers letters of a word. It is also considered how to structure these words in the paragraph. There are three main methods which are used in the language approach.

#### A. Bag of Words (BOW):

There is an equal importance of each in the paragraph and each word is considered as independent entities. Analysis of individual words are performed and signs of misinformation is detected. By investigating patterns, misleading information can be detected.

#### B. Semantic Analysis:

The truthfulness of information can be identified using personal experience. An honest writer makes similar remarks on the topic. Different compatibility approach can be used to find out truthfulness of information.

**C. Deep Syntax:** This method is using probability context free grammar for analysis [18]. Sentences are converted into rewritten rules for analysing various syntax structures. These sentences are compared with already known structures and patterns of lies that tells differences between real news and fake news [19]

#### 2. Approach of Topic-agnostic

In this approach, the fake news detection is considering topic-agnostic features and capabilities of web markup to identify the fake news. This approach does not content of article for identifying the fake news [20]. Following are topic-agnostic approaches:

- (i) A huge advertisement
- (ii) Headlines of longer type that has phrases of eye-catching.
- (iii) Various text patterns in mainstream news for emotive responses
- (iv) Using author name

#### 3. Approach of Machine Learning

This approach is using different type of datasets to train the algorithm. Datasets are also used to develop new machine algorithms for detecting the fake news. There are some domains for which datasets can be collected. These domains are sports, business, entertainment, politics, technology and education. Fig. 2 presents various approaches of machine languages which can be used to detect fake news.

A machine learning approach [21] which is called rumour identification is to identify ambiguous post as fake news. This modal is built for twitter [7] and it has four main areas:

- (i) Metadata for tweets,
- (ii) Origin of the tweet
- (iii) Date of the tweet and Area of the tweet,
- (iv) Location of the tweet

#### 4. Approach of Knowledge Based

This approach integrated approach of machine learning and knowledge base for detecting the fake news. The challenge in detecting the fake news is the speed of fake news spreading on the social media. One example of rapid spread of small piece of information is on twitter (Qazvinian et al. 2011).

Knowledge-based system is using the external resources to identify fake news very quickly before spreading it. Following are three categories [13]:

- (i) Fact checking with help of expert

- (ii) Fact checking using computation
- (iii) Fact checking using crowd sourcing

(i) Fact checking with help of expert  
 In this approach, data is analyzed and examine carefully. It requires professionals to evaluate news manually by checking the accuracy of text with another text has been checked previously.

(ii) Fact checking using computation  
 It uses a tool which automatically check the facts for identifying a news is true or false. It uses open web sources and knowledge graphs for differentiating between fake news and real news. A tool called ClaimBuster for automatically identifying fake news. This tool uses natural language processing and machine learning techniques. The tool examines context of the text in speeches on social media in real time to find out facts and compares it to read facts.

(iv) Fact checking using crowd sourcing  
 This approach provides the opportunity to a group to take the collective decision. The group examines the accuracy of the news. The accuracy is completely depending on wisdom of the group or cloud. This group first examines piece of information then the group looks for entire sentence for identifying its truthiness.

**4.5 Approach of Hybrid**

Three elements in fake news article:

- (i) text of the article
- (ii) response that article received
- (iii) source used to motivate that article

The approach uses a hybrid model a combination of machine learning and human being and for identifying the fake news. Humans are able to identify the fake news 54% of time [14]. The hybrid model increases this percentage. The effectiveness of hybrid model combines a network approach and social media news with machine learning. The aim of this hybrid model is to find probability about fake news. One more hybrid model is discussed here called CSI (Capture, Score, integrate). This model functions on the following events:

- (1) Capture –  
 Extracting representation of articles using Recurrent Neural Network (RNN)
- (2) Score: It creates a score and representation vector
- (3) Integrate: It is used to integrate the outputs of the capture and score resulting in a vector which can be used for classification.

**IV. METHODOLOGY**

To detect the fake news, it is to use classification algorithms. Following following machine learning algorithms are ap-

plied XGBoost, Liner Regression, Random Forest, K-Nearest Neighbors (K-NN), Decision Tree, Naïve Bayes, and Support Vector Machine (SVM). This model detects the political fake news by applying natural language tool kit. The training data is selected for 80% and remaining is used for 20% dataset.

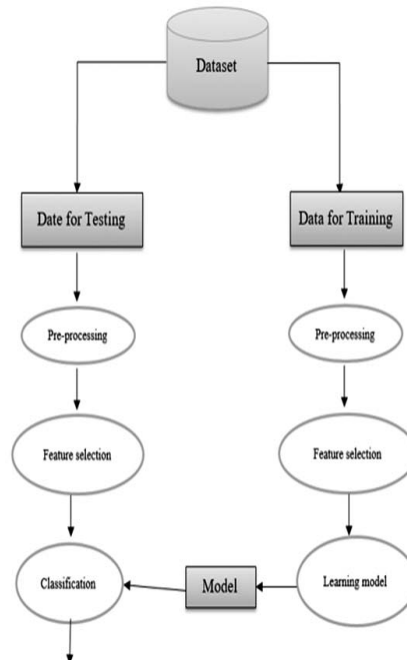


Fig.1: Methodology [12]

There are three main fake news contributors: social bots, trolls, and cyborg users. A social bot is able to generate content and interact with users automatically. Trolls are humans who creates fake news for online communities for provoking users into an emotional response for social media.

Cyborg is a mixture of automated activities with the help of human input. Cyborg users are able to switch between functionalities between human and bot. URLs are collected by crawling the web for finding clickbaits in social media website. The job of crawler is to collect tweets and stores in the database. Users are interested to know the credibility of a new tweet.

Fake news consists of misleading stories with no verification [14]. It is motivated to direct towards reputation damage and economic gain. Fake new is not new. It exists since beginning of print journalism but it has been spreading fast because of IT development and impacting to large people.

The research shows that untrustworthy sites have been increased by 305 percent between 2019 and 2020 [16] and credible information has been increased by just 69%. Fake news has also shown result in economic losses [17].

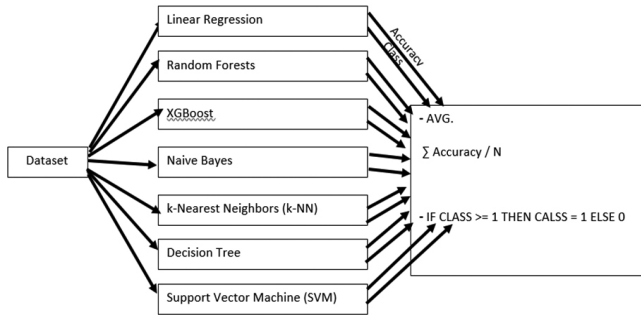


Fig. 2: Various algorithms [12]

Focus on fake news across a range of political issues. Health-related fake news creates pandemic situation. Similarly, war-related fake news also affects greatly at international level [16]. Sometimes technology related fake news is also disseminated that impacts on the economy of the county [15]. Use of online political fake news is from long time to gain political benefit.

Over last few years, people are using online political fake news for political gain. Sometimes, misinformation has positive effect on the electoral support for populist parties.

## REFERENCES

- [1] Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, 141, 215-222.
- [2] Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying fake news and fake users on Twitter. *Procedia Computer Science*, 126, 451-461.
- [3] Geeng, C., Yee, S., & Roesner, F. (2020, April). Fake news on Facebook and Twitter: Investigating how people (don't) investigate. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-14).
- [4] Aldwairi, M., Hasan, M., Balbahaith, Z., 2017b. Detection of drive-by download attacks using machine learning approach. *Int. J. Inf. Sec. Priv.* 11, 16–28.
- [5] Agarwal B, Xie I, Vovsha O, Rambow R, Passonneau R. Sentiment Analysis of Twitter Data. In *Proceedings of the ACL 2011, Workshop on Languages in Social Media*, 2011, p. 30-38.
- [6] Barbosa L, Feng J. Robust Sentiment Detection on Twitter from Biased and Noisy Data. *COLING 2010*, p. 36-44.
- [7] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter. *WWW 2011 – Session: Information credibility*, ACM, 1, Hyderabad, India, 2011, p. 675-684.
- [8] Shellenbarger S. Most Students Don't Know When News Is Fake, Stanford Study Finds. *The Wall Street Journal*, 21 November 2016: <https://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576> accessed last time on March, 2018.
- [9] Pierson D. Facebook and Google pledged to stop fake news. So why did they promote Las Vegas-shooting hoaxes? *Los Angeles Time*, 2 October 2017: <http://www.latimes.com/business/la-fi-tn-vegas-fake-news-20171002-story.html> accessed last time on March, 2018.

- [10] Clark CS. Fake news? A survey on video news releases and their implications on journalistic ethics, integrity, independence, professionalism, credibility, and commercialization of broadcast news. PhD Thesis, College of Communication and Information Sciences in the Graduate School of the University of Alabama, 2009.
- [11] <https://www.coe.int/en/web/campaignfree-to-speak-safe-to-learn/dealing-withpropaganda-misinformation-and-fake-news>
- [12] Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021, March). Fake news detection using machine learning approaches. In *IOP conference series: materials science and engineering* (Vol. 1099, No. 1, p. 012040). IOP Publishing.
- [13] Ahmed, S., Hinkelmann, K., & Corradini, F. (2022). Combining machine learning with knowledge engineering to detect fake news in social networks-a survey. *arXiv preprint arXiv:2201.08032*.
- [14] Okoro, E.M., Abara, B.A., Umagba, A.O., Ajonye, A.A., Isa, Z.S.: A hybrid approach to fake news detection on social media. *Niger. J. Technol.* 37(2), 454 (2018). <https://doi.org/10.4314/njt.v37i2.22>
- [15] Desai, S., Mooney, H., & Oehrli, J. A. (2021). Fake News, Lies and Propaganda: How to Sort Fact from Fiction. *Subjects: News & Current Events*. The University of Michigan Library.
- [16] Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., & de Figueiredo Nicolete, L. D. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review. *Journal of Public Health*, 1-10.
- [17] Velichety, S., & Shrivastava, U. (2022). Quantifying the impacts of online fake news on the equity value of social media platforms—Evidence from Twitter. *International Journal of Information Management*, 64, 102474.
- [18] Stahl, K. (2018). Fake news detection in social media. *California State University Stanislaus*, 6, 4-15.
- [19] Burkhardt, J. M. (2017). History of fake news. *Library Technology Reports*, 53(8), 5-9.
- [20] Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., & Freire, J. (2019, May). A topic-agnostic approach for identifying fake news pages. In *Companion proceedings of the 2019 World Wide Web conference* (pp. 975-980).
- [21] De Beer, D., & Mathee, M. (2021). Approaches to identify fake news: a systematic literature review. *Integrated Science in Digital Age 2020*, 13-22.



**Dr Shashank Sahu** received his B.E. in Computer from North Maharashtra University, Jalgaon, India in 1996. He received his M.Tech degree in Computer Science & Engineering from Gautam Buddha Technical University, Lucknow, India, in 2010. He received his Ph.D. degree in Computer Science & Engineering from Sharda University, Greater Noida, India in June 2020. He is working as Professor in Computer Science & Engineering Department at Ajay Kumar Garg Engineering College Ghaziabad, India. He has 24 years of academic experience. He is the author of more than 45 publications in national/international journals and conferences and guided 12 M.Tech. students. His research areas are Software Engineering, Web Technologies, Programming Paradigm, Intelligent Agent and Object-Oriented Programming.