IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 9, September 2020

DOI 10.17148/IJARCCE.2020.9917

A Review of Fake News Detection using Machine Learning

Atul Suryawanshi[1], Himanshu Yadav[2], Chetan Agrawal[3]

M.Tech Scholar, Department of Computer Science Engineering,

Radharaman Institute of Technology & Science, Bhopal, India[1]

Assistant Professor, Department of Computer Science Engineering,

Radharaman Institute of Technology & Science, Bhopal, India[2]

Associate Professor & HOD, Department of Computer Science Engineering,

Radharaman Institute of Technology & Science, Bhopal, India[3]

Abstract: Now days the usage of Internet and online marketing has become very popular. Millions of products and services are available in online marketing that generate huge amount of information. Social media sites can have a major influence in expanding the span of this kind of story. Fake news is a news created to intentionally misguide or mislead readers. Fake news is spread mainly for gaining political or financial incentives. There has been a large surge of fake news in recent times due to the immense use of social media and online news media. It has become much easier to spread fake news then how it used to be earlier. This kind of fake news when spread may have a severe effect. Hence it is extremely essential that certain measures should be taken in order to reduce or distinguish between real and fake news. This paper presents a survey on fake news detection based on various supervised, unsupervised and semi supervised data mining and machine learning techniques.

Keywords: Machine Learning, Fake News, supervised, unsupervised and semi supervised.

I. INTRODUCTION

News is basic part of our life. In regular daily existence current news are valuable to improve data what happen all over. So most of society favor watching news by far most of the society all around slant toward examining paper quickly in the initial segment of the day getting an accuse out of cup of tea. If news is phony that will trick social orders occasionally counterfeit word used to get out gossipy goodies about things or it will impact some political pioneer positions taking into account counterfeit news. By far most of the serious cell phone customers need to scrutinize the news by methods for online media over web.

The news destinations are circulating the news and give the wellspring of affirmation. The request is the best approach to check the news and articles which are flowed among electronic media like WhatsApp social occasions, Facebook Pages, Twitter and other scaled down scale online diaries and individual to individual correspondence objections. It is hazardous for the overall population to acknowledge on the pieces of chatter and affirm to be a news. So it's crucial to find the phony news. Online news organizes phenomenally sway our overall population and culture in both positive and negative ways.

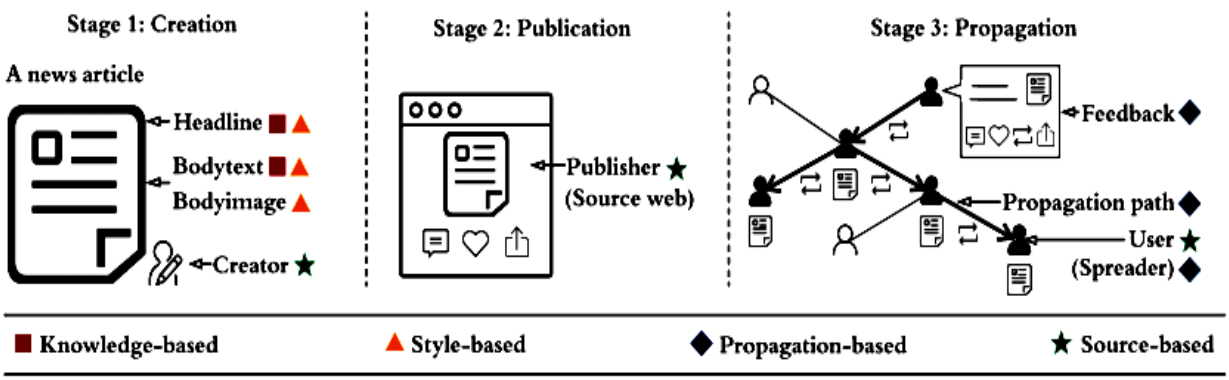

Figure 1: Fake News Life Cycle

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 9, September 2020
DOI 10.17148/IJARCCE.2020.9917

As online media ends up being more dependent for wellspring of information, a lot of phony news is posted on the web, that wide with people following it with no prior or complete information of event validity. Such misdirection can control famous sentiments. The exponential improvements of phony news inciting have become a remarkable peril to open for news steadfastness. It has become a persuading issue for which finding, inspecting and overseeing counterfeit news has extended famous. Spreading of deception on the web these days addresses a troublesome issue, as their impact on people's appraisals may be immense.

Fake news news addresses a specific kind of double dealing. While its location was by and large being performed genuinely previously, automated procedures using AI and related fields ended up being more fundamental. On the other hand, significant learning methods ended up being standard and a great part of the time used techniques in the field of information examination starting late.
This audit paper presents a survey of the phony news recognition and discussion of the promising assessment heading. The key motivations of this examination are summarized as follows:

- Fake news through online media has been going on for a serious drawn-out period of time; nevertheless, there is no unending gracefully of the term counterfeit news". To all the almost certain guide the future headings of phony news recognition research, appropriate clarification are fundamental.
- Social media has wind up being a mind boggling hotspot for counterfeit news spread. There are some rising models that can be utilized for counterfeit news identification in online media. A review on existing phony news recognition methods under various online media circumstances can give a basic understanding on the top tier counterfeit news discovery procedures.

Fake news discovery by means of online media is still in the early season of headway, and there are up 'til now many testing issues that need further assessments. It is critical to discuss potential investigation heading that can improve counterfeit news discovery and easing limits.

II. LITERATURE SURVEY

A. Uppal et al., propose and execute a strategy for modernized misleading recognition. The proposed system uses significant learning in talk level structure examination to detail the structure that isolates phony and real news. The standard model achieved 74% precision[1].
V. M. Krešňáková, and C. K. Hiramath et al., introduced the location of phony news from the printed information utilizing profound learning methods. thought was to prepare various kinds of neural organization models utilizing both whole messages from the articles and to utilize only the title text. The models were prepared and assessed on the Phony News dataset acquired from the Kaggle rivalry. Likewise proposed counterfeit news discovery framework dependent on characterization, for example, Strategic relapse (LR), Innocent bayes (NB), Backing vector machine (SVM), Arbitrary woods (RF) and profound neural organization (DNN) and think about all AI strategies for distinguishing counterfeit news [2][3].
I. Kareem et al. presents the recognizable proof of phony news, in view of 344 News stories and named it physically Phony or Valid, explored two component extraction procedures like Term Recurrence (TF) and Term Recurrence Backwards Report Recurrence (TF-IDF). Seven diverse regulated AI (ML) grouping calculations are utilized and their outcomes examination has done. Best execution classifier K Closest Neighbors (KNN) gives 70% precision and strategic relapse gives 69% accuracy[6].
K. Rajesh et al., presents a classifier that can anticipate whether a bit of news is genuine and not only a messed up truth. The proposed model train itself utilizing informational collections having features of updates on numerous years to anticipate whether a news story is consistent with its promise. The proposed work gives a helpful issue free stage for everybody and expects to spread quiet by diminishing bits of gossip and misconceptions in the general public [7]. B. M. Amine et al., presents a combined profound learning model that recognize counterfeit articles with respect to various qualities. Subsequently, we use word installing strategy and convolutional neural organization to remove text based highlights and analyze diverse design of profound learning while at the same time blending two CNNs with various metadata (Text, title, and creator). We show on genuine dataset that the proposed approach is exceptionally proficient and permits to accomplish high performances[8].
H. Telang et al., presents the issue from an information situated viewpoint by exploring whether programmed computational methodologies in NLP and AI can be utilized to distinguish lies in composed content. Execution of highlights like n-grams and word vectors utilized with five managed learning strategies in recognizing Counterfeit News stories are analyzed. The effect of specific changes in the boundaries of highlight extraction on classifier execution are likewise broke down [9].
I. AYDIN et al., presents AI based techniques were utilized to identify counterfeit records that could misdirect individuals. For this reason, the dataset created was pre-prepared and counterfeit records were dictated by AI calculations. Choice trees, strategic relapse and backing vector machines calculations are utilized for the recognition of phony records.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 9, September 2020
DOI 10.17148/IJARCCE.2020.9917

Grouping exhibitions of these techniques are contrasted and the strategic relapse demonstrated with be more fruitful than the others [10].

R. K. Kaliyar et al., first present the datasets which contain both phony and genuine news and direct different investigations to compose counterfeit news finder. At that point utilize Normal Language Preparing, AI and profound learning procedures to group the datasets [11]. A. Kathal et al., presents approach to find the best model to predict the popularity of online news using machine learning methods. Initially, correlation techniques are used to gain dependence on the popularity received from an article and to obtain attributes or characteristics that are optimal for subsequent classification. [12].

During the literature reviews some point observed that is following-

- The parameters like precision, recall, Fmeasure, accuracy, error rate and time is calculated by the researchers.
- No efficient approach to optimize the good values of these parameters.
- Simulation time is also a key parameter to judge the proposed algorithm approach.
- Low accuracy rate and high error rate makes failure and inefficient of the algorithm.
- Less sample size data sets taken by existing research papers.

III. EVALUATION METRICS

To evaluate the performance of algorithms for fake news detection problem, various evaluation metrics have been used. In this subsection, we review the most widely used metrics for fake news detection. Most existing approaches consider the fake news problem as a clarification problem that predicts whether a news article is fake or not:

- True Positive (TP): when predicted fake news pieces are actually annotated as fake news
- True Negative (TN): when predicted true news pieces are actually annotated as true news
- False Negative (FN): when predicted true news pieces are actually annotated as fake news
- False Positive (FP): when predicted fake news pieces are actually annotated as true news

By formulating this as a clarification problem, we can define following metrics,

$$Precision = \frac{|TP|}{|TP| + |FP|}$$

$$Recall = \frac{|TP|}{|TP| + |FN|}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

IV. DETECTION TECHNIQUES

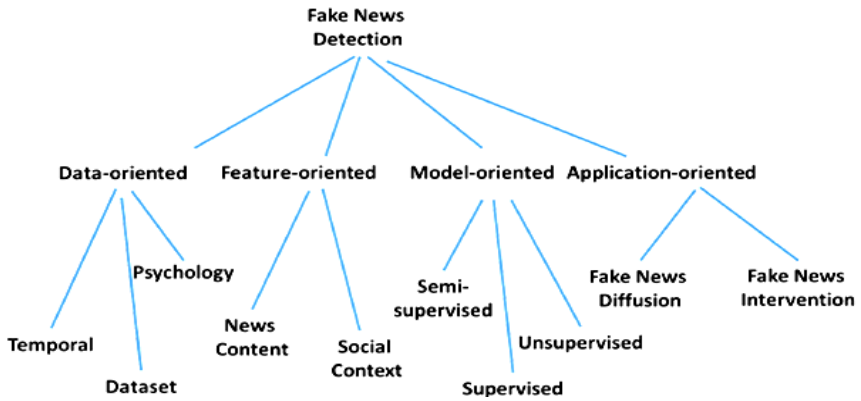

Figure 2: Future directions and open issues for fake news detection on social media

Various techniques have been proposed in past to identify fake reviews based on types of data like labelled data (for example, supervised learning), unlabeled data (for example, unsupervised learning), and partially labeled data (for example, semi-supervised learning) that is described below.

A. Supervised Learning techniques

Before applying the classification method, different preprocessing steps are performed; these steps include stemming, removal of punctuation marks and stop word removal. They use linguistic feature to identify fake reviews. Linguistic feature contains POS and bag-of-words. Bag-of-words features consist of individual word or group of words that are

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 9, September 2020
DOI 10.17148/IJARCCE.2020.9917

found in given text. Then different classification algorithms are applied like decision tree, random forest, support vector machine, naive bayes and gradient boosted trees. Here naive bayes and support vector machine give better result.

B. Unsupervised Learning techniques

Main advantage of unsupervised learning approach is that, without any labeled dataset, we can classify fake and genuine reviews. This concept uses different features based on review data, reviewer data and product information based on difference in behavioral pattern of reviews. Here author uses Amazon cell phone reviews dataset to identity fake and genuine reviews.

C. Semi-Supervised Learning techniques

Positive Unlabeled (PU) learning technique is combination of some positive label and unlabelled dataset. PU-learning technique is semi supervised technique, which only uses two class classifiers positive as deceptive and unlabeled without having negative as truthful training example. In this algorithm, firs unlabeled data are considered as negative class. In next step, classifiers are trained based on initial set of positive instances. Then classifiers are applied only on unlabeled instances and generate labeled instances. After, classified positive and negative instances, the positive instances as deceptive reviews are eliminated from unlabeled instances and rest of them are considered as negative instances. Again classifiers are applied into negative instances. This process is repeated until the stop criteria, which classify fake and genuine reviews. Here two classifiers are applied in PU learning, support vector machine and naive bayes.

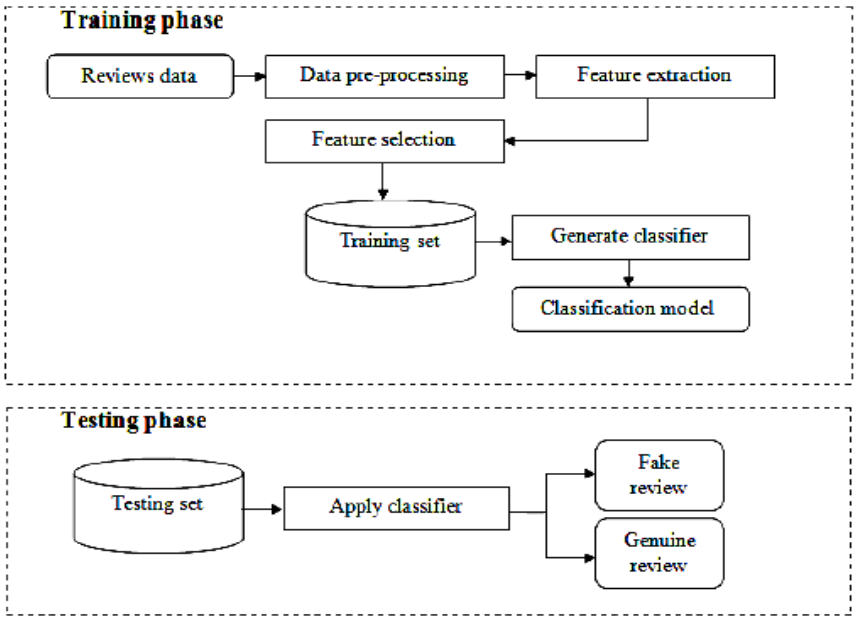

Figure 3: Machine Learning based Fake Review Detection

There are different approaches to distinguish counterfeit surveys. AI strategy is one of the approaches to distinguish counterfeit audits [10]. AI model learns and make forecast [2]. The fundamental advances associated with AI are information handling, highlight extraction, include determination, order model age. This cycle is appeared in Fig. 3: AI approach for counterfeit survey location functions as follows:

- **Data collection:** In this stage, survey information will be assembled from different stages like Amazon. These audits could be for item or administration like inn surveys.
- **Data pre-processing:** In subsequent stage, information preprocessing is applied like accentuation marks expulsion, stemming, stop word evacuation and so forth. In accentuation marks evacuation, the entire content is isolated into sentences, expressions or sections. In the stemming cycle, stem will be made from each word in dataset. In stop word evacuation stage, as often as possible utilized gathering of words like determiners, articles and relational word will be distinguished and taken out. Subsequent to eliminating these words, just significant words will be held for the following stage.
- **Classifier model construction and testing:** For preparing reason, little arrangement of named information is utilized. In this stage, grouping model is created by utilizing the preparation audit dataset. The surveys utilized for this design are as of now marked as phony or veritable audit. When the classifier is prepared, it will be tried utilizing test dataset. The distinctive AI calculations which can be utilized for model development are guileless bayes order, choice tree calculation, uphold vector machine, k-closest neighbor, strategic relapse, and so forth. The exhibition of phony survey identification strategy relies upon named information utilized for preparing reason, right determination of highlights and information digging procedures utilized for recognition.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 9, September 2020

DOI 10.17148/IJARCCE.2020.9917

V. CONCLUSION

Machine Learning uses a statistical technique to give the computer the ability to learn with data hence it is widely used in the detection of fake news. Methods used for taking parameters and for categorizing the type of news are also discussed. With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has strong negative impacts on individual users and broader society. From the literature review it has been observed that the accuracy for predicting fake news in social media is much higher than any other online news media hence we have targeted online news media fake news detection along with website verification. In this article, we review the fake news detection based on the data mining and the machine learning approach. In future work, our proposed model will be tested for fake news detection by using standard dataset and apply machine learning based algorithm for detection and validation. Our approach will improve the performance parameters and make a efficient algorithm.

REFERENCES

[1]. A. Uppal, V. Sachdeva and S. Sharma, "Fake news detection using discourse segment structure analysis," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 751-756, doi: 10.1109/Confluence47617.2020.9058106.

[2]. V. M. Krešňáková, M. Sarnovský and P. Butka, "Deep learning methods for Fake News detection," 2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTI-MACRo), Szeged, Hungary, 2019, pp. 000143-000148, doi: 10.1109/CINTI-MACRo49179.2019.9105317.

[3]. C. K. Hiramath and G. C. Deshpande, "Fake News Detection Using Deep Learning Techniques," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 411-415, doi: 10.1109/ICAIT47043.2019.8987258.

[4]. A. Jain, A. Shakya, H. Khatter and A. K. Gupta, "A smart System for Fake News Detection Using Machine Learning," 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), GHAZIABAD, India, 2019, pp. 1-4, doi: 10.1109/ICICT46931.2019.8977659.

[5]. R. K. Kaliyar, A. Goswami and P. Narang, "Multiclass Fake News Detection using Ensemble Machine Learning," 2019 IEEE 9th International Conference on Advanced Computing (IACC), Tiruchirappalli, India, 2019, pp. 103-107, doi: 10.1109/IACC48062.2019.8971579.

[6]. I. Kareem and S. M. Awan, "Pakistani Media Fake News Classification using Machine Learning Classifiers," 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 2019, pp. 1-6, doi: 10.1109/ICIC48496.2019.8966734.

[7]. K. Rajesh, A. Kumar and R. Kadu, "Fraudulent News Detection using Machine Learning Approaches," 2019 Global Conference for Advancement in Technology (GCAT), BANGALURU, India, 2019, pp. 1-5, doi: 10.1109/GCAT47503.2019.8978436.

[8]. B. M. Amine, A. Drif and S. Giordano, "Merging deep learning model for fake news detection," 2019 International Conference on Advanced Electrical Engineering (ICAEE), Algiers, Algeria, 2019, pp. 1-4, doi: 10.1109/ICAEE47123.2019.9015097.

[9]. H. Telang, S. More, Y. Modi and L. Kurup, "Anempirical Analysis of Classification Models for Detection of Fake News Articles," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869504.

[10]. İ. AYDIN, M. SEVİ and M. U. SALUR, "Detection of Fake Twitter Accounts with Machine Learning Algorithms," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-4, doi: 10.1109/IDAP.2018.8620830.

[11]. R. K. Kaliyar, "Fake News Detection Using A Deep Neural Network," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-7, doi: 10.1109/CCAA.2018.8777343.

[12]. A. Kathal and M. Namdev, "Correlation Enhanced Machine Learning Approach based Online News Popularity Prediction", IJOSCIENCE, vol. 4, no. 3, p. 6, Mar. 2018. https://doi.org/10.24113/ijoscience.v4i3.124