Fake News Detection Through ML and Deep Learning Approaches for Better Accuracy

Anil Kumar Dubey and Mala Saraswat

Abstract Social media form a very frequent podium to people to freely express their opinions and easily communicate to others. Nowadays, it plays a vital role for spreading the news headlines, and it became most applicable news sources globally as easily accessible, but also risky as exposure of "fake news" misleads the people. The extensive spread of such misinformation deploys negative impacts on people and society and becomes recently a global problem. Several issues already rise in worlds during elections process, due to huge spread of fake news. Therefore, the detection of it on social platform transforms into an emerging research that is exciting enormous concern. Problem to identifying the fake news has concentration to public as well as government organization. Such propaganda probably affects the opinion of people and malicious parties involved to manipulate the conclusion. Due to the majority of society opinion impact changes, fake news detection is an important challenge to researchers. The detection of misinformation is not an easy task for anyone, but quite is a complex for people. Here, we analyze the different fake news detection approaches followed in current scenario and compute the detection process through machine learning and deep leaning algorithms for better accuracy.

Keywords News · Social media · Detection · Classification · Accuracy

1 Introduction

Several misinformation spread on society through social media and lead to change the opinion of people. The detection of this distinguishes from fact in it and is very important task for society. Researchers of several areas are investigating the mounting production as well as diffusion of misinformation that rapidly infects the society. "The news that is purposely and certainly false and is able to mislead the readers" is known

A. K. Dubey (✉) · M. Saraswat
ABES Engineering College Ghaziabad, Ghaziabad, Uttar Pradesh 201009, India

M. Saraswat
e-mail: mala.saraswat@abes.ac.in

as fake news [1]. According to Merriam Webster Online Dictionary, fake news is "the news reports that are intentionally false or misleading" [2]. Generally, we can say that fake news is unauthenticated information that circulates or propagated on any platform for misleading to human brain, and it includes rumors, propaganda, and satire [3].

Basic

The mounting curiosity to detect the fake news fascinated to researchers due to the circulation of fake information through social media platforms. Such misinforming content is quickly spreads on social media and gains popularity. As people received the content from social media, they easily can believe on it and interpret their minds for it as reliable source of information to get trust. Due to blind trust and easily acceptance of people for fake news information, several solemn and unenthusiastic impacted fake articles get viral and observe on society and people leads to a disparity of news ecosystem [4]. Nowadays, social media (i.e., Facebook, WhatsApp, and Twitter) have covered most of the developing countries worldwide to become the major resource for news. It is essential in economics, social developments, and politics for motivated human brain and influences the process of these in negative impact and finally targets to damage to public figures and agencies [5].

Several active researchers are enduring to detect the fake news from social media. The detection process is estimating the misleading information of any news/topic whether it is planned or accidentally affected to society. In most of the cases, the detection process deploys the machine learning algorithm to filter the news, whether it is fake or not [5]. The detection of such things is very difficult task due to slight differences between fake and real news [6]. Current scenario of messaging shows that the fake news has become an ordinary thing for it. No one can refute the contents easily as it may fake; still their meaning depends on view of person [7]. The identification of fake news is an imperative subject for both public and society. The amplifying social media interaction has led the increment in number of people to unfair [8].

Problem

From past few decades, social media cover globe as they become the major resource of news information, due to the accessibility, minimum cost. Other perspective illustrates the danger to expose "fake news" aimed to mislead and manipulate the readers. The extensive spread of such misinformation deploys negative impacts on people and society and becomes recently a global problem. One major issue rises in U.S 2016 presidential elections due to huge spread of fake news. Therefore, the detection of it on social platform transforms into an emerging research that is exciting enormous concern.

Objective

The main objective of this work is accurate detection of fake news using machine learning algorithms.

2 Related Work

Researchers' enduring valuable contribution in this era is to distinguish the fake news from facts. Kaur et al. study and evaluate the superior supervised algorithm to detect the fake news. Author concerns about to superior learning classifier of machine learning on the basis of diverse conditions for detection, and appropriate model that gets best detection in particular condition is elaborated. Due to fake news getting viral on social media related to Covid-19 will impact society because our prediction will be based on data set collected from fake viral news. Similarly, a fake video was spread few years before on social media related to Kerala battling for floods, and it became so much viral. The news claims that the Chief Minister of Kerala state was forced to Indian Army to stop the rescue operations in state of flooded regions. One of the famous fake news was viral through WhatsApp groups in India in 2019 during national election was impacted to India's ruling party [4]. Harita Reddy analyzed several approaches to detect the fake news through text features. Author gets the 95.49% accuracy in detection through combo of stylometric and text features [9]. Nicollas et al. elaborate the analysis to detect the fake news through text extraction of social media, using the natural language processing. Author uses news data from Twitter as 33,000 tweet and distinguished the real and fake news among them. Approx. 86% accuracy is received through dimension reduction of original features [10]. Rubin et al. discuss about the details of fake news as stated that it can be divided into three parts as: pure fraudulent nature target to confuse the readers, rumors, and sarcasm and irony [11].

According to Peining Shi et al., malicious social bots are spreading the misinformation to mislead the society, therefore, wish to detect and remove these bots from social networks. Generally, this detection uses the easily imitated quantitative features for behavioral analysis and receives the low accuracy of detection. Author presents the joint approach as transition probability-based feature selection and semi-supervised clustering for detection [12]. Ghafari et al. discuss the trust concept for social networks and trust-related challenges to prediction process. Author classifies the trust prediction through addressing the challenges and invites the contributor for this era [13]. Day to day the emerging technologies arise, a need of viral reduction methodology is acquired for fake news to control the misleading of society. Shrivastava et al. present a model to evaluate the fake news propagation and describe how fake news spreads among several groups. Authors considers the current pandemic as COVID-19 for viral fake news [14]. Umer et al. discuss a fake news detection stance model based on headlines and news body. Author used the principal component analysis, chi-square for quality features extraction and also concerns to dimensional reduction approach for better result. PCA is used for noise removal and discusses model gain approximately 97.8% accuracy [15].

Domenico and Visentin discuss the marketing-related fake news and studied their details as consumer behavior, marketing ethics, future avenue, and strategy for fake news from eighty-six scientific articles and five managerial reports [1]. Ajao et al. present the fake news characteristics related to sentiments and process of

fake news detection. Author analyzes the text-based fake news detection considering both included and excused sentiments on Twitter dataset [3]. Elhadad elaborate the systematic survey of fake news detection on social media till 2017. Author discussed different types of fake news and presents general overview of summarization of news documents with different features that are extracted from news. Author notices that as spreads of fake news in social media, the detection system is not sufficient and its shortage invites researchers for more contribution in this era. Several prospectives are vacant for detailed contribution in big data of fake news [5]. Kuai Xu et al. highlight the continuous growth in fake news on social media that impacted the society. Authors target to analyze the differences between real and fake news based on their status and domain uniqueness. Kuai et al. used neural network for distinguishing the text in high-dimensional vector space for analysis [6]. Wenlin Han and Varshil Mehta discuss and evaluate the performance to detect the fake news in social networks through machine learning as well as deep learning algorithm. The fake news spreads rapidly in society leading to misguide the opinion of people, due to the fasted and easiest medium to transmit the information. The misleading information creates major impact on reader's brain for manipulated aspects. Authors use naïve Bayes, hybrid convocational neural network, recurrent neural network algorithm for it [16]. Hanz and Kingsland discuss about a news that is real or fake in details. As in presidential election 2016 at the USA, it created lot of information to mislead the people and impacted their brain. One workshop was organized to discuss the hole and flaw of viral information for election and analyzed the tweets to compute the reality and compared from previous [17]. Rajesh et al. discussed about a classifier to predict reality in viral news slice. Authors used the several years' news headlines to compare and for prediction process of news reality through natural language processing to mine the text [7]. Correia et al. focused to detect the fake news with new feature extraction, and analysis for practical application, and also concern about offers as well as challenges for it [18]. Day to day, the fake news identification becomes most popular issues for society, due to growth of social network users. Vereshchaka et al. stated that the fake news becomes the issues of not only for individuals but also as societal issues due to continuous people growing interaction on social media and technical challenges to distinguish the fake and reality of news. As per statistics of research, more than two millions of users deleted every month by famous social media as WhatsApp to stop the spread of the misleading information [8].

As several researchers already contributed in this era of fake news detection, but still some more efforts are required for detection as day to day grow-up in the social media users, so researcher is continuing to work in this era for more accurate and advance detection of fake news.

3 Proposed Approach

We collect data from Kaggle and preprocess the data for missing and unwanted data. After preprocessing, the different ML algorithms will perform one by one and check the accuracy of fake news detection. The algorithm which gains higher accuracy is pointed out. Decision tree algorithm and XGBoost provide the best accuracy in prediction of fake news. Author also applied the long short-term memory (LSTM) algorithm to predict the higher accuracy for ideal condition of acceptance.

Data Preprocessing

Before applying the classification techniques of decision tree, it is required to preprocess the data, for a definite alteration as shuffle, stop word and punctuation removal from text, grouping, lower casing, word clouds, and tokenization. The preprocessing process optimizes the data as per requirements from original size. The general preprocessing techniques are used to remove punctuation and non-letter typescript; after that, the lowered casting is performed. In addition, word cloud is used to represent the words in graphical way, and tokenization is done to count number of tokenized data frame. Stop words are irrelevant words normally used in sentence for their structure formation and generate the noise during classification. These words are removed from original data, and processed data is stored for next step (Fig. 1).

Fig. 1 Proposed approach for fake news detection
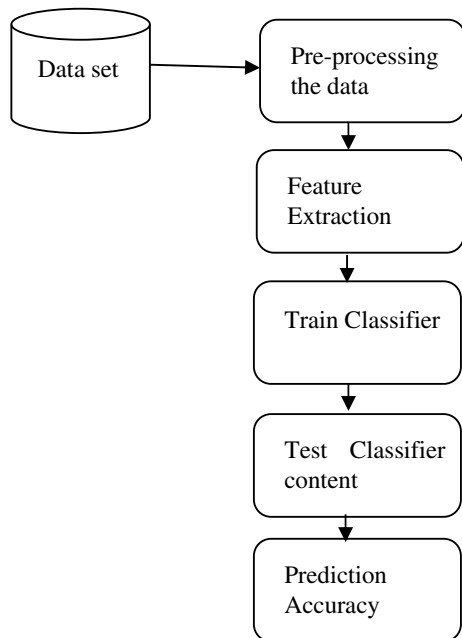
Features extraction

Several terms, phrases, and words may present in the data that show the extra load for computational to the learning process, and also some irrelevant features impact the classifier performance and accuracy. Therefore, its feature reduction is very important task that reduces the features size in feature space dimension.

Train Classifier

Select the appropriate classifier as decision tree for classification and split the data into two parts as training and testing. The target plan for classifier training is up to eighty percent of text data using random state.

Test Classifier

After training process of text data, the testing phase continues with target plan up to twenty percent of text data using random state. Prediction Accuracy. The decision tree is most popular technique for prediction and classification. The decision tree classifier accuracy for false news prediction will be computed with considering parameters.

4 Experiments

In this section, we demonstrate the detection through decision tree as best machine learning algorithm for prediction with Kaggle datasets for accuracy of detection. After that, we applied XGBoost algorithm and LSTM algorithm for better accuracy [19, 20]. In considered data, the shape is (44,898, 5) defined, in which (23,481, 4) fake news and (21,417, 4) true news shape ordered.

Figure 2 illustrates the graphical representation of word cloud data in database for fake and real news. The first part of figure as (a) represents the fake news world cloud; similarly, (b) represents true news world cloud. This operation performs through preprocessing of data, in which the world counted and word text-based cloud are formed. Decision tree is the supervised machine learning algorithm for continuously
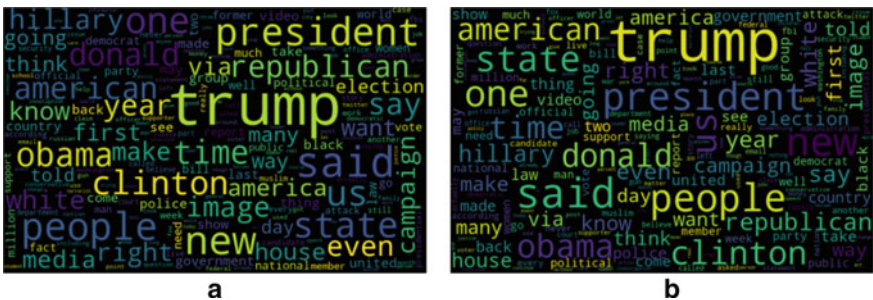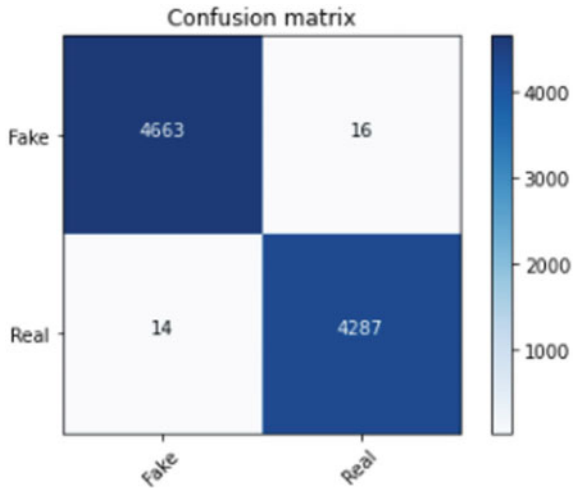

Fig. 2 **a** Fake word cloud, **b** true word cloud

Fig. 3 Confusion matrix of decision tree for fake news detection

splitting data based on certain parameter. This classifier is used to divide and conquer approach to split data into subsets and again subsets as required. Therefore, author considers this algorithm for prediction of fake news. By applying the vectorizing the text in pipeline of vector count with maximum depth of tree considered as twenty and random state up to forty-two for transformer, gain of the confusion matrix is as illustrated in Fig. 3.

After applying the decision tree algorithm, ninety-nine point six seven (99.67%) percent for prediction of fake news is captures. After successfully applied the decision tree, author processes for best accuracy and applies decision tree-based ensemble ML algorithm (XGBoost) where a gradient boosting framework is used. Author is also applying the LSTM on dataset. LSTM is the type of recurrent neural network that has the capability to learn order dependency in sequence prediction problems. After processing both, author compared the result as mentioned in Table 1.

After applying the XGBoost algorithm, ninety-nine point seven (99.7%) percent for prediction of fake news is captures, that is, greater than the decision tree algorithm. Finally, the accuracy ninety-nine point nine (99.9%) percent for prediction of fake news is captures from long short-term memory. As it is near about to ideal condition of prediction as hundred percent, author did not check another algorithm for prediction.

Table 1 Comparison of performance of ML and deep learning approach for fake news detection

Method	Accuracy	Precision	Recall	F1-score
XGBoost	0.997	0.85	0.85	0.85
LSTM	0.999	0.87	0.87	0.87

5   Conclusion

The digital age of technology motivated people to interact with social media for news and messages. Due to high interaction of population of society, people post, transfer, and gain the news as well as messages from this. And, some illegal group is disturbing this phenomenon of accepting the news via posting the illegal information. As the human brain mostly faiths on it, it cannot distinguish the viral fake news and accept the viral news as real news and society as well as individual's brain is changed for that. Therefore, it is very important task for organization to control such rumors to spread from society and also detect the fake news. Author contributed in this era is targeting to accuracy for better prediction. As decision tree classifier is providing the best solution in most of the cases of prediction, therefore, applying this algorithm is to predict the fake news and gain most acceptable accuracy of prediction. After that, author also applied XGBoost and LSTM algorithms for better accuracy and received ideal condition of acceptance to accuracy. In the future, author plans to use different type of complex data and big data of fake news for classification to be capture with the best accuracy.

References

1. Domenico GD, Visentin M (2020) Fake news or true lies? Reflections about problematic contents in marketing. Int J Mark Res 62(4):409–417. https://doi.org/10.1177/147078532093 4719
2. Fake news—political scandal words. [Online]. Available https://www.merriam-webster.com/words-at-play/politicalscandal-words/fake-news
3. Ajao O, Bhowmik D, Zargari S (2019) Sentiment aware fake news detection on online social networks. In: ICASSP 2019. 978-1-5386-4658-8/18/$31.00 ©2019 IEEE, pp 2507–2511
4. Kaur S, Kumar P, Kumaraguru P (2019) Automating fake news detection system usingmulti-level voting model. SpringerVerlag GmbH Germany, part of Springer Nature 2019. https://doi.org/10.1007/s00500-019-04436-y
5. Elhadad MK, Li KF, Gebali F (2019) Fake news detection on social media: a systematic survey. 978-1-7281-2794-1/19/$31.00 ©2019 IEEE
6. Xu K, Wang F, Wang H, Yang B (2020) Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Sci Technol 25(1):20–27. ISSN 1007-0214 03/14. https://doi.org/10.26599/TST.2018.9010139
7. Rajesh K, Kumar A, Kadu R (2019) Fraudulent news detection using machine learning approaches. In: 2019 global conference for advancement in technology, India. 978-1-7281-3694.3/19/$31.00 ©2019 IEEE
8. Vereshchaka A, Cosimini S, Dong W (2020) Analyzing and distinguishing fake and real news to mitigate the problem of disinformation. In: Computational and mathematical organization theory. S.I.: SBP-BRIMS 2019, © Springer Science+Business Media, LLC, part of Springer Nature 2020.https://doi.org/10.1007/s10588-020-09307-8

9. Reddy H, Raj N, Gala M, Basava A (2020) Text-mining-based Fake News Detection Using Ensemble Methods. IJAC, © Institute of Automation, Chinese Academy of Sciences and Springer-Verlag GmbH Germany, part of Springer Nature 2020. https://doi.org/10.1007/s11633-019-1216-5

10. de Oliveira NR, Medeiros DSV, Mattos DMF (2020) A sensitive stylistic approach to identify fake news on social networking. IEEE Sig Process Lett. https://doi.org/10.1109/LSP.2020.3008087

11. Rubin VL, Chen Y, Conroy NJ (2015) Deception detection for news: three types of fakes. In: 78th ASIS&T annual meeting: information science with impact: research in and for the community. American Society for Information Science, p 83

12. Shi P, Zhang Z, Kwang K, Choo R (2019) Detecting malicious social bots based on clickstream sequences. IEEE Access.https://doi.org/10.1109/ACCESS.2019.2901864

13. Ghafari SM, Beheshti A, Joshi A, Paris C, Mahmood A, Yakhchi S, Orgun MA (2020) A survey on trust prediction in online social networks. IEEE Access. https://doi.org/10.1109/ACCESS.2020.3009445

14. Shrivastava G, Kumar P, Ojha RP, Srivastava PK, Mohan S, Srivastava G (2020) Defensive modeling of fake news through online social networks. IEEE Trans Comput Soc Sys. © IEEE 2020. https://doi.org/10.1109/TCSS.2020.3014135

15. Umer M, Imtiaz Z, Ullah S, Mehmood A, Choi GS, On BW (2016) Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access. https://doi.org/10.1109/ACCESS.2017

16. Han W, Mehta V (2019) Fake news detection in social networks using machine learning and deep learning: performance evaluation. In: 2019 IEEE ICII. 978-1-7281-2977-8/19/$31.00 ©2019 IEEE. https://doi.org/10.1109/ICII.2019.00070

17. Hanz K, Kingsland ES (2020) Fake or for real? A fake news workshop. Ref Serv Rev 48(1):91–112. © Emerald Publishing Limited, 0090-7324. https://doi.org/10.1108/RSR-09-2019-0064

18. Reis JCS, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. Affective computing and sentiment analysis. IEEE Intell Syst. 1541-1672_2019 IEEE, Published by the IEEE Computer Society. https://doi.org/10.1109/MIS.2019.2899143

19. Sahoo SR, Gupta BB (2021) Multiple features based approach for automatic fake news detection on social networks using deep learning. Appl Soft Comput 100:106983

20. Choudhary M, Chouhan SS, Pilli ES, Vipparthi SK (2021) BerConvoNet: a deep learning framework for fake news classification. Appl Soft Comput 110:107614

21. Jindal R, Dahiya D, Sinha D, Garg A (2022) A study of machine learning techniques for fake news detection and suggestion of an ensemble model. In: International conference on innovative computing and communications. Springer, Singapore, pp 627–637

22. Sharma DK, Garg S (2021) IFND: a benchmark dataset for fake news detection. Complex Intell Syst 1–21

23. Zervopoulos A, Alvanou AG, Bezas K, Papamichail A, Maragoudakis M, Kermanidis K (2021) Deep learning for fake news detection on Twitter regarding the 2019 Hong Kong protests. Neural Comput Appl 1–14

24. Kaliyar RK, Goswami A, Narang P (2021) FakeBERT: fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools Appl 80(8):11765–11788

25. Khanam Z, Alwasel BN, Sirafi H, Rashid M (2021, March) Fake news detection using machine learning approaches. IOP Conf Ser Mater Sci Eng 1099(1):012040. IOP Publishing

26. Divya TV, Banik BG (2021) A walk through various paradigms for fake news detection on social media. In: Proceedings of international conference on computational intelligence and data engineering. Springer, Singapore, pp 173–183

27. Dubey AK, Singhal A, Gupta S (2020) Rumor detection system using machine learning. Int Res J Eng Technol (IRJET) 07(05). e-ISSN 2395-0056

28. Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ (2021) An ensemble machine learning approach through effective feature extraction to classify fake news. Futur Gener Comput Syst 117:47–58