

E-ISSN: 2707-6644

P-ISSN: 2707-6636

IJCPDM 2023; 4(1): 62-64

Received: 11-01-2023

Accepted: 15-02-2023

Dr. Naveeta Adlakha
Assistant Professor,
Department of Computer
Science, Govt. College for
Girls, Palwal, Kurukshetra,
Haryana, India

A comparative analysis of machine learning algorithms for fake news detection

Dr. Naveeta Adlakha

DOI: <https://doi.org/10.33545/27076636.2023.v4.i1.a.81>

Abstract

With the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. In this paper, aim is to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real. Fake news detection is an emerging research area which is ahead big interest. It faces however some challenges due to the limited resources such as datasets and processing and analysing techniques.

Keywords: Internet, social media, fake news, artificial intelligence, websites, authenticity

Introduction

The fake news can be spread for different drives. Some are made only to increase the number of clicks and visitors on a particular website while others, to influence public opinion on political choices or on economic markets. Fake news concerning health on social media represents a risk to global worth.

Now a days a large amount of our time is spent interacting online through social media platforms for latest news updates rather than traditional news organizations^[1]. the reality behind the scene is that (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism, like newspapers or television; and (ii) it's easier to further share, discuss, and discuss the news with friends or other readers on social media. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. The extensive spread of faux news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem. Fake news is typically manipulated by propagandists to convey political messages or influence.

Also the fake news changes the way people interpret and answer real news. To assist mitigate the negative effects caused by fake news (both to profit the general public and therefore the news ecosystem). It's crucial that we build up methods to automatically detect fake news broadcast on social media.

Internet and social media have made the access to the news information much easier and comfortable^[2].

Characteristics of Fake News

They often have grammatical mistakes. They are often emotionally coloured. They often try to affect readers' opinion on some topics. Their content is not always true. They often use attention seeking words and news format and click baits. They are too good to be true. Their sources are not genuine most of the times^[9].

Literature Survey

Mykhailo Granik *et al.* in their paper^[3] shows a simple approach for fake news detection using naive Bayes classifier.

Corresponding Author:
Dr. Naveeta Adlakha
Assistant Professor,
Department of Computer
Science, Govt. College for
Girls, Palwal, Kurukshetra,
Haryana, India

This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

Himank Gupta *et al.* [9] gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (BotMaker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150,000 spam tweets and 250,000 non-spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of-Words model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately 18%.

Marco L. Della Vedova *et al.* [10] first proposed a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing its accuracy up to 78.8%. Second, they implemented their method within a Facebook Messenger Chabot and validate it with a real-world application, obtaining a fake news detection accuracy of 81.7%. Their goal was to classify a news item as reliable or fake; they first described the datasets they used for their test, then presented the content-based approach they implemented and the method they proposed to combine it with a social-based approach available in the literature. The resulting dataset is composed of 15,500 posts, coming from 32 pages (14 conspiracy pages, 18 scientific pages), with more than 2,300,000 likes by 900,000+ users. 8,923 (57.6%) posts are hoaxes and 6,577 (42.4%) are non-hoaxes.

Cody Buntain *et al.* [11] develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets: CREDBANK, a crowd-sourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumours in Twitter and journalistic assessments of their accuracies. They apply this method to Twitter content sourced from BuzzFeed's fake news dataset. A feature analysis identifies features that are most predictive for crowd-sourced and journalistic accuracy assessments, results of which are consistent with prior work. They rely on identifying highly re-tweeted threads of conversation and use the features of these threads to classify stories, limiting this work's applicability only to the set of popular tweets. Since the majority of tweets are rarely re-tweeted, this method therefore is only usable on a minority of Twitter conversation threads.

Comparison of Machine Learning Algorithms for Fake News Detection

Many works are interested to fake news detection.

- Amy *et al.* [2] proposed a typology of several methods of truth valuation emerging from two main categories: linguistic cue methods with machine learning and network analysis approaches, for detecting fake news.
- Granik *et al.* [3] presented a simple approach to fake news detection using a naive Bayesian classifier. This approach is tested on a set of data extracted from

Facebook news posts. They claim to be able to attain an accuracy of 74%. The rate of this model is good but not the best, as many other works have achieved a better rate using other classifiers. That We discuss these works in the following.

- Kai Shu [1] propose a fake news detection model that uses n-gram analysis and machine learning techniques by comparing two different feature abstraction techniques and six different classification methods or techniques. The experiments carried out show that the best presentations are obtained by using the so-called features extraction method (TF-IDF). They used the Linear Support Vector Machine (LSVM) classifier that gives an accuracy of 92%. This model uses LSVM that is limited to treat only the case of two linearly separated classes.
- Yang [14] shows that how users of social networks can guarantee the truth and reliability of information. They also describe the mechanisms that allow their authentication and the role of journalists or what to expect from researchers and official institutions. This work assists people see a little bit of the truth behind the news on social media and not believe anything blindly.
- Metz [5] offer several strategies and types of indices relating to different sense modality (text, image, social information). They also discover the value of combining and merging these approaches using AI to assess and verify shared information.
- Nair [8] *et al.* present an overall performance analysis of different approaches on three different datasets. This work focused on the text of the info. And the feeling given by it, and ignores some features like the source, the author or the date of the journal that can have a dramatic impact on the result.
- Conroy [6] created a new public dataset of valid new articles and proposed a text-processing based machine learning approach for automatic documentation of Fake News with 87% accuracy. It seems that this work focuses on the emerging feelings from the text and not on the content of the text in itself.
- Yang [14] introduced LIAR, a new dataset for automatic fake news detection. This quantity can also be used for stance classification, argument mining, topic modeling, rumor detection, and political NLP research. Most of the works in this area have used this benchmark. However, it is well-known that this last is restricted to political information, while others have integrated info. From various fields of the resources.

Drawbacks in Existing Algorithms

The general disadvantage of these approaches is that the unconditional data encoding may not be valid in reality. Besides, the usual fake news organization is limited to two values (i.e., namely, Real or Fake), while in reality we can't say that the news is real or fake at 100%, but according to a degree of confidence. We consider that this point is very important to classify the news in social media.

Future Work

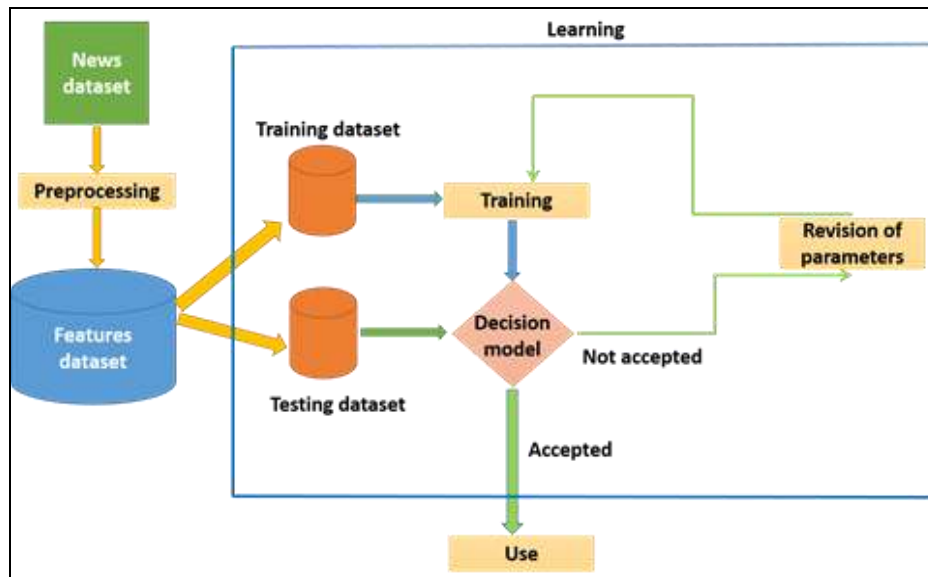
An innovative method is required for detecting fake news. That can use a news dataset to build a decision model based on support vector machine method. Then the model can be used to classify novel news to fake or real.

This can be achieved by

- **Text preprocessing:** Consisting of roasting and analyzing the text by removing stop words and special characters.
- **Encoding of the text:** Using bag of words and N-gram then TF-IDF.
- **Extraction of the features:** It allows an accurate

documentation of false data. We use the source of a news, its author, the date and the feeling given by the text as features of a news.

- **Support vector machine:** A supervised machine learning algorithm that allows the classification of new information.

**References**

1. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu. —Fake News Detection on Social Media: A Data Mining Perspective arXiv:1708.01967v3 [cs.SI], 3 Sep, 2017
2. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu, —Fake News Detection on Social Media: A Data Mining erspectiveI arXiv:1708.01967v3 [cs.SI], 3 Sep, 2017
3. Granik M, Mesyura V., Fake news detection using naive Bayes classifier, 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev; c2017. p. 900-903.
4. Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website.
5. Accessed Feb. 6, 2017
6. Cade Metz. The bittersweet sweepstakes to build an AI that destroys fake news, Dec. 16, 2016.
7. Conroy N, Rubin V, Chen Y. Automatic deception detection: Methods for finding fake newsI at Proceedings of the Asso. for Information Scienc & Technology. 2015;52(1):1-4.
8. Markines B, Cattuto C, Menczer F. Social spam detectionI. In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web; c2009. p. 41-48.
9. Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, —Fake News Detection using Machine Learning and Natural Language Processing,I International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, 2019, 7(6).
10. Gupta H, Jamal MS, Madisetty S, Desarkar MS. A framework for real-time spam detection in Twitter, 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru; c2018. p. 380-383
11. Della Vedova ML, Tacchini E, Moret S, Ballarin G, DiPierro M, de Alfaro L. Automatic Online Fake News Detection Combining Content and Social Signals, 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla; c2018. p. 272-279.
12. Buntain C, Golbeck J. Automatically Identifying Fake News in Popular Twitter Threads, 2017 IEEE International Conference on Smart Cloud (Smart Cloud), New York, NY; c2017. p. 208-215.
13. Parikh SB, Atrey PK. Media-Rich Fake News Detection: A Survey, 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL; c2018. p. 436-441
14. Scikit-Learn- Machine Learning In Python Dataset-Fake News detection William Yang Wang. liar, liar pants on _re: A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648; c2017.
15. Shankar Patil M, Dr. Praveen Kumar. —Data mining model for effective data analysis of higher education students using MapReduceI IJERMT, April 2017, 6(4).