

# False Data Prediction On Social Media Using Machine Learning

**Pranita Balwadkar**

Department of Computer Engineering, Siddhant College of Engineering, Sudumbare

**Abstract:** Internet is one of the most significant inventions, and many people utilise it. These people employ it for various functions. These users have access to a variety of social networking channels. Through these internet platforms, every user can submit something or share a story. These sites don't check the content that users publish, so some individuals try to propagate false information on them. These false reports may serve as propaganda against a specific person, group, company, or political party. A person cannot discern all of these false reports. Therefore, there is a need for machine learning classifiers that can automatically identify these phoney news. This comprehensive literature review describes the use of machine learning classifiers for detecting false news.

**Keywords:** Online fake news, text classification, machine learning, fake news, social media

## I. INTRODUCTION

The world is rapidly evolving. This digital world certainly has some benefits for us, but it also has some drawbacks. In this digital world, there are numerous issues. Fake news is one of them. False information can easily spread. A person or organization's reputation is harmed by spreading false information. It could be propaganda directed at a person or an organization, such as a political party. There are different online platforms where the person can spread the fake news. This includes Twitter, Facebook, and others. According to Donepudi (2019), machine learning is a component of artificial intelligence that aids in the creation of systems that are capable of learning and carrying out various actions. The supervised, unsupervised, and reinforcement machine learning algorithms are just a few of the many machine learning algorithms that are available. The algorithms must first be trained using a dataset known as the train data set. These algorithms can be used for a variety of tasks following training. Machine learning is using in different sectors open form different tasks. The majority of the time, machine learning algorithms are used to predict or discover hidden information.

Users benefit from online platforms because they can easily access news. But the problem is that this gives cybercriminals the chance to spread fake news on these platforms. This news can be proved harmful to a person or society. After reading the news, readers begin to believe it without verifying it. Due to the difficulty of the task, identifying fake news presents a significant obstacle (Shuetal). (2017). If fake news is not caught early, people will spread it to others, and everyone will start to believe it. Through fake news, people, groups, and political parties can be affected. In the 2016 US election, fake news affected people's opinions and decisions (Dewey, 2016).

Various researchers are working to identify fake news. In this regard, machine learning is proving to be beneficial. Different algorithms are being used by researchers to find false news. According to researchers (Wang, 2017), it is difficult to identify fake news. They have detected fake news using machine learning. Zhou et al. researchers, (2019) found that fake news spreads more and more as time goes on. Because of this, fake news detection is necessary. This is what the algorithms of machine learning are trained to do. Once they have been trained, machine learning algorithms will automatically recognize fake news.

The various research questions will be answered by this literature review. The significance of AI to recognize counterfeit news will be demonstrated in this writing survey. The application of machine learning to the detection of false news will also be the subject of discussion. The literature review will talk about machine learning algorithms used to find fake news.

The rest of the paper has the following structure: methodology in section 2, research questions in section 3, search process model in section 4, literature review results and discussion in section 5, and conclusion in section 6. The papers that are the subject of this literature review are cited in the final section.

**II. METHODOLOGY**

This writing survey is composed for responding to some exploration questions. Therefore, the systematic literature review was used as the methodology. The research questions can be answered with this method. The papers that will be discussed in this literature review were gathered from various databases. To respond to the examination questions, different exploration papers are talked about and referred to in this literature review.

**Exclusion and Inclusion**

Every day, a few papers are published. Therefore, a number of papers are displayed in the result when a string is searched. Not all of the papers pertain to that string. This indicates that the criteria are required. The table below lists the inclusion and exclusion criteria used in this literature review.

Exclusion Criteria	Inclusion Criteria
The language of the paper is not the English language.	Papers that are written in the English language.
The complete paper is not accessible.	Paper can be accessed completely.
Paper is not related to machine learning and fake or false news detection.	Paper showing content related to machine learning and fake or false news detection.

Table 1: Exclusion and Inclusion Criteria

The literature review included papers that met the a fore mentioned inclusion criteria.

**Quality Assessment**

The research work presented in each of the included papers was used to evaluate the quality of the papers. The papers in which the researchers discussed the application of machine learning to the detection of fake or false news were deemed to be of high enough quality to be included in this literature review.

**III. RESEARCH QUESTION**

Some RQs must be answered by ASLR. On the basis of sound arguments, this literature review will provide responses to three research questions. The following is a list of the two research questions.

RQ1: Why is it necessary to use machine learning to identify fake news?

RQ2: Which supervised classifiers for machine learning can be used to identify fake news?

RQ3: How are machine learning classifiers trained to recognize fake news?

The results and discussion section of this literature review will provide answers to these research questions.

**IV. SEARCH PROCESS**

To find the papers that can be discussed in this literature review, a search process is used. The diagram provided below makes it simple to comprehend this search procedure.

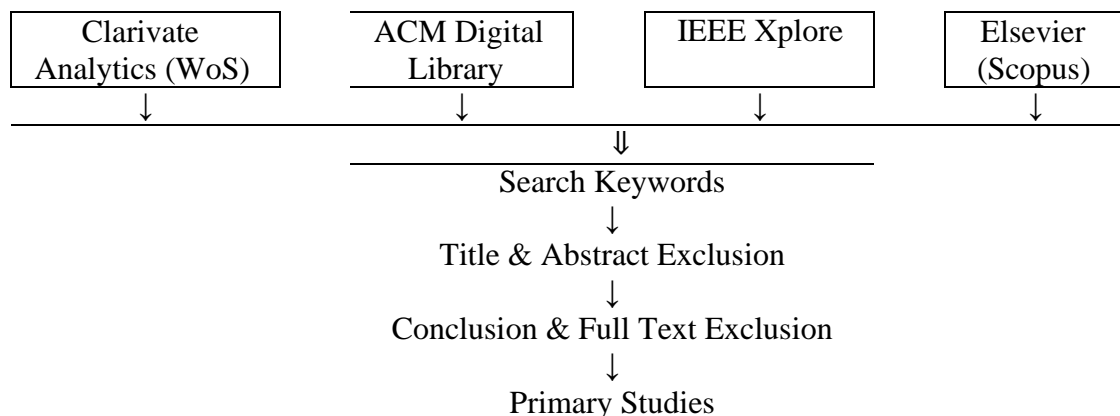


Figure 1: Search Process Model Diagram

The papers came from a variety of databases. However, not all of them were pertinent to the subject. Therefore, the papers were first excluded based on their titles and abstracts. An abstract is a brief summary of the paper's contents that can give an idea of the paper's content. The remaining papers were examined against the inclusion and exclusion criteria in the subsequent phase. Seventy-three papers were retrieved from various databases in response to the search keyword. Twenty-six of the papers discussed in this literature review remained after the exclusion.

## V. RESULT AND DISCUSSION

According to Donepudi (2020), the internet is one of the best sources of information for its users. Social media platforms like Facebook and Twitter make it easier for people to connect with others. On these platforms, different kinds of news are also shared. These days, people prefer to read the news on these platforms because they are simple to use and accessible. The availability of comment, response, and other options on these platforms is yet another benefit to the public. People use these platforms because of these benefits (Donepudi et al., 2020b). However, these platforms are also used by cybercriminals as the best source, just like their advantages. These individuals can use these platforms to spread false information. On these platforms, there is also a feature for sharing posts or news, which helps spread fake news. As a result, people begin to believe the news and spread it to others. Zubiaga et al. researchers, (2018) stated that it is challenging to prevent the spread of false information on these social media platforms.

On these platforms, anyone can sign up and start spreading news. A person can create a news source page and spread false information. These platforms do not validate the individual's credibility as a publisher. Anyone can use this method to spread negative information about a person or organization. A society or political party can also be hurt by these false news. According to the report (Levin, 2017), it is simple to change people's opinions by spreading fake news. As a result, it is necessary to prevent the spread of fake news in order to preserve a person's, a political party's, or an organization's reputation.

RQ1: Why is it necessary to use machine learning to identify fake news?

The spread of false information has become easier as more people use the internet. Fake news can be disseminated to a large number of people through a variety of social media platforms. The news quickly spread thanks to these platforms' share option. False news has an impact not only on individuals but also on organizations and businesses (Donepudi et al., 2020a). Therefore, controlling fake news is necessary. A person can only tell if a piece of news is fake if he or she knows the whole story. The majority of people don't know the whole story, so they just start believing fake news without verifying it. This makes it hard to do.

Because no one can control fake news, the question of how to control it arises here. Machine learning is the solution. Fake news can be detected with the assistance of machine learning (Khan et al., 2019). These fake news can be easily and automatically detected using machine learning (Della Vedova et al., 2018). Machine learning algorithms will examine the content of the post and determine that it is fake news once it has been posted by someone. According to Krusinski (2020), a number of researchers are searching for the most effective machine learning classifier for identifying fake news. The classifier's accuracy must be taken into account because if it fails to identify fake news, it could be harmful to various individuals. The training of this classifier is what determines its accuracy. Accuracy can be improved by a well-trained model. The following question will provide an answer to a variety of machine learning classifiers that can be used to identify fake news.

RQ 2: Which supervised classifiers for machine learning can be used to identify fake news?

One of humankind's most difficult tasks is to identify fake news. Machine learning makes it simple to identify fake news. Various machine learning classifiers can assist in determining whether a piece of news is true or false. It is now simple to collect the dataset necessary to train these classifiers. For the purpose of verifying the veracity of news, various researchers utilized machine learning classifiers. Abdullah-All-Tanvir et al. researchers, (2019) detected fake news using machine learning classifiers. The SVM and Naive Bayes classifiers, as demonstrated by the researchers' experiments, perform the best at identifying fake news. Based on their accuracy, these two classifiers perform better than others. A more accurate classifier is thought to be superior. The most important thing is how accurate any classifier is. More accurate classifiers will aid in the detection of more fake news. According to researchers (Kudavalli & Fiaidhi, 2020), it is important to catch fake news because many people use social media to spread false information. To prevent individuals and businesses from losing their reputations as a result of false news, it is necessary to detect it (Rahman et al.). (2020). According to them, machine learning is extremely beneficial in this regard. Using a variety of machine learning algorithms, they discovered that logistic regression is a superior classifier due to its higher accuracy.

According to Aphiwongsophon & Chongstitvatana (2018), researchers, social media generate a large number of posts. Any post can be made by anyone who registers on these platforms. False information about a person or business could be in this post. It is a challenging and important task to identify such false news. Researchers have utilized the three machine learning techniques for this task. The neural network, neural network, and SVM are these. The neural network and SVM, on the other hand, achieved an accuracy of 90.90%, while the Native Bayes method achieved an accuracy of 96.08%.

According to there searchers of(Ahmedetal. 2017), false news has a significant impact on a society's political situation. People's perceptions can be altered by false news posted on social media platforms. Fake news causes people to shift their perspectives without verifying them. A method that can detect such news is required. For this purpose, the researchers have utilized machine learning classifiers. The K-Nearest Neighbor, Support Vector Machine, Logistic Regression, Linear Support Vector Machine, Decision Tree, and Stochastic Gradient Descent are some of the classifiers utilized by various researchers. The results indicate that the linear support vector machine performed well in identifying false news.

Researchers (Reiset al. (2019) have utilized machine learning classifiers to identify fake news. These classifiers have been trained with a variety of features. Because a trained classifier can produce more accurate results, training the classifiers is an essential task. Artificial intelligence, according to researchers (Granik&Mesyura, 2017), is superior at identifying fake news. They have identified fake news in Facebook posts using a Naive Bayes classifier. Their accuracy with this classifier was 74%, but they said it could be better. In that paper, these researchers also describe various methods for increasing accuracy. Fake news can be identified using machine learning classifiers.

Some of these popular classifiers are given below that are used for this purpose.

**Support Vector Machine:** Classification is the primary application for this algorithm. This is a labeled data set-based supervised machine learning algorithm. Scientists in (Singh et al. , 2017) used a variety of machine learning classifiers and a support vector machine to find fake news the best.

**Naive Bayes:** The classification tasks also make use of Naive Bayes. You can use this to determine whether the news is real or fake. Pratti et al. researchers, 2017) identified the fake news using this machine learning classifier.

**Logistic Regression:** When the value that needs to be predicted is categorical, this classifier is used. It can, for instance, predict or indicate whether the result is true or false. In the (Kaur et al. ,2020) have utilized this classifier to determine whether a piece of news is genuine or fake.

**RandomForests:** There are various random forests in this classifier that each assign a value, and the value with the most votes is the actual result. Ni et al., 2020) researchers have identified fake news using a variety of machine learning classifiers. The randomforest is just one of these classifiers.

**Recurrent Neural Network:** Additionally, this classifier is useful for identifying fake news. The recurrent neural network has been used by researchers in (Jadhav & Thepade, 2019) to determine whether news is true or false.

**Neural Network:** Classification problems can benefit from the use of various machine learning algorithms. Neural networks are one of these algorithms. Researchers from the (Kaliyar et al., 2020) have detected fake news using a neural network.\

**K-Nearest Neighbor:** Classification problems can be solved with this supervised machine learning algorithm. This stores information about all cases in order to classify the new case based on similarities. The scientists (Kesarwani et al., 2020) have detected fake news on social media using this classifier.

**Decision Tree:** This supervised algorithm of machine learning can help to detect the fake news. It divides the data set into various smaller subsets. Researcher sin(Kotteti et al. (2018)) have utilized a variety of machine learning classifiers, including the decision tree. They have identified the fake news by employing these classifiers.

**RQ3: How machine learning classifiers are trained for detecting fake news?**

The task of training machine learning classifiers is crucial. This has a significant impact on the accuracy of these classifiers' results. A classifier must be properly trained using the appropriate data set. The machine learning classifiers have been trained by a variety of researchers to recognize fake news. The majority of the training data set is in an imbalanced form, which is the primary issue that arises when training these classifiers (Wang et al., 2020). Al Asaad and Erascu (2018) used supervised machine learning classifications to identify fake news. The three distinct models for feature extraction were utilized in order to train these classifiers. In fact, the classifiers are trained with these features. The TF-IDF Model, the N-Gram Model, and the Bag of Words Model are these models. The features are extracted from the training data set by these models, and the classifier is trained using these features. A group of researchers (Ahmed et al., 2018) has taught a few machine learning classifiers to spot fake news. They used a training data set for the purpose of training. The words have been reduced to their singular form after the redundant ones have been removed first. so that only useful data should be included in the classifier's training dataset.

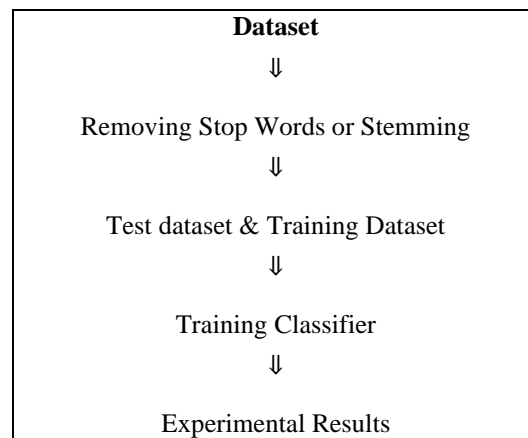


Figure 2: TrainingDataset

Figure 3 shows the steps that are used while training a classifier. After the training a classifier is then used for experiments.

## VI. CONCLUSION

It is now simple to spread fake news because more people use the internet. A large number of people regularly use social media and the internet. When publishing news on these platforms, there are no restrictions. As a result, some individuals take advantage of these platforms and begin disseminating fake news against individuals and organizations. This can harm an individual's reputation or a company. Through counterfeit news, the assessments of individuals can likewise be changed for an ideological group. There must be a method for identifying these false news. Classifiers based on machine learning are used for a variety of things, and they can also be used to find fake news. A data set known as the training data set is used to first train the classifiers. After that, these classifiers are able to recognize fake news on their own.

The supervised machine learning classifiers that require labeled data for training are discussed in this systematic literature review. It is difficult to find labeled data that can be used to train classifiers for detecting fake news. The application of unsupervised machine learning classifiers to the detection of fake news could be the subject of future research

## REFERENCES

- [1] Abdullah-All-Tanvir, Mahir, E. M., Akhter S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp.1-5, <https://doi.org/10.1109/ICSCC.2019.8843612>
- [2] Ahmed,H.,Traore,I.,&Saad,S.(2017).Detectionofonlinefakenewsusingn-gramanalysisand machine learning techniques. Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 127–138, Springer, Vancouver, Canada, 2017. [https://doi.org/10.1007/978-3-319-69155-8\\_9](https://doi.org/10.1007/978-3-319-69155-8_9)
- [3] Ahmed, H., Traoré, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. Secur.



- Priv., 1(1), 1-15. <https://doi.org/10.1002/spy2.9>
- [4] Al Asaad, B., & Erascu, M. (2018). A Tool for Fake News Detection. 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2018, pp.379-386. <https://doi.org/10.1109/SYNASC.2018.00064>
- [5] Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting Fake News with Machine Learning Method. 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 528-531. <https://doi.org/10.1109/ECTICon.2018.8620051>
- [6] Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L. (2018). Automatic online fake news detection combining content and social signals. FRUCT'22: Proceedings of the 22st Conference of Open Innovations Association FRUCT. Pages 272-279. <https://dl.acm.org/doi/10.5555/3266365.3266403>
- [7] Dewey, C. (2016). Facebook has repeatedly trended fake news since firing its human editors. The Washington Post, Oct. 12, 2016.
- [8] Donepudi, P. K. (2019). Automation and Machine Learning in Transforming the Financial Industry. Asian Business Review, 9(3), 129-138. <https://doi.org/10.18034/abr.v9i3.494>
- [9] Donepudi, P. K. (2020). Crowdsourced Software Testing: A Timely Opportunity. Engineering International, 8(1), 25-30. <https://doi.org/10.18034/ei.v8i1.491>
- [10] Donepudi, P. K., Ahmed, A. A. A., Saha, S. (2020a). Emerging Market Economy (EME) and Artificial Intelligence (AI): Consequences for the Future of Jobs. Palarch's Journal of Archaeology of Egypt/Egyptology, 17(6), 5562-5574. <https://archives.palarch.nl/index.php/jae/article/view/1829>
- [11] Jadhav, S. S., & Thepade, S. D. (2019). Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier, Applied Artificial Intelligence, 33(12), 1058-1068. <https://doi.org/10.1080/08839514.2019.1661579>
- [12] Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet—A deep convolutional neural network for fake news detection. Cognitive Systems Research, 61, 32-44. <https://doi.org/10.1016/j.cogsys.2019.12.005>
- [13] Kaur, S., Kumar, P. & Kumaraguru, P. (2020). Automating fake news detection system using multi-level voting model. Soft Computing, 24(12), 9049-9069. <https://doi.org/10.1007/s00500-019-04436-y>
- [14] Kesarwani, A., Chauhan, S. S., & Nair, A. R. (2020). Fake News Detection on Social Media using K-Nearest Neighbor Classifier. 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, pp.1-4. <https://doi.org/10.1109/ICACCE49060.2020.9154997>
- [15] Khan, J. Y., Khondaker, M., Islam, T., Iqbal, A., & Afroz, S. (2019). A benchmark study on machine learning methods for fake news detection. Computation and Language. <https://arxiv.org/abs/1905.04749>
- [16] Kotteti, C. M. M., Dong, X., Li, N., & Qian, L. (2018). Fake news detection enhancement with data imputation. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). Athens, 2018, pp.187-192. <https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00042>
- [17] Kudarvalli, H. & Fiaidhi, J. (2020). Experiment on Detecting Fake News using Machine Learning Algorithms. International Journal of Reliable Information and Assurance, 8(1), 15-26. <https://dx.doi.org/10.21742/IJRIA.2020.8.1.03>
- [18] Kurasinski, L. (2020). Machine Learning explainability in text classification for Fake News detection. Malmö University Publications. <https://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-20058>
- [19] Levin, S. (2017). Pay to sway: report reveals how easy it is to manipulate elections with fake news. The Guardian. 13 Jun 2017
- [20] Ni, B., Guo, Z., Li, J., & Jiang, M. (2020). Improving Generalizability of Fake News Detection Methods using Propensity Score Matching. Social and Information Networks. <https://arxiv.org/abs/2002.00838>