IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Fake Social Media Profile Detection And Reporting Using Machine Learning Algorithms

Pathan Asma , Chitra Gayathri , Appireddygari Vijetha, Golla Anusha Sai ,
Ms. Amirtha Preeya V
Presidency school of computer science and Engineering, Presidency University, India.

ABSTRACT- Technology is advancing rapidly every need for an effective tool that can accurately detect fake accounts.
Classification algorithm is used to identify these fake accounts. Fake news is a term that can have different meanings to different people. At its core, fake news can be defined as fabricated and without enough sources, verifiable facts, or quotes. Researchers discovered that individuals are increasingly likely to encounter false and fabricated information in their daily life. Some surveys state that manipulative cascades are spreading between the ratio of 1000 to 100,000 people whereas if we talk about the true information then it barely reaches 1000 people. With respect to this research problem, we also came to know that politicians and stock marketers use these types of practices to achieve their agenda, or we can say people generally use such methods to get their work done, make profits, or gain power.

1. INTRODUCTION

Social media has touched everyone's life as number of people on social media is expanding exponentially. Instagram has seen a great increase and got prominence among web-based social accounts. It is most famous internet-based platform, but also used for online frauds, spreading fake information through social media at a rapid pace. There is a widespread need for an effective tool that can accurately detect

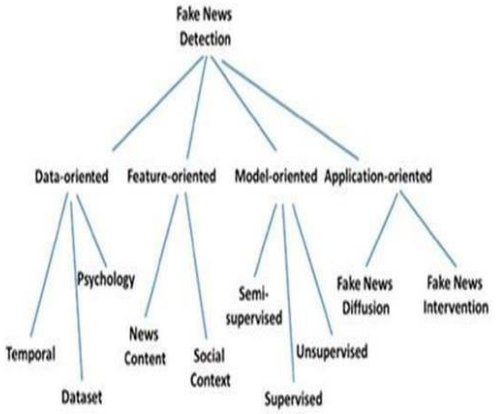
fake accounts. Classification algorithm is used to identify these fake accounts. Fake news is a term that can have different meanings to different people. At its core, fake news can be defined as fabricated and without enough sources, verifiable facts, or quotes. Researchers discovered that individuals are increasingly likely to encounter false and fabricated information in their daily life. Some surveys state that manipulative cascades are spreading between the ratio of 1000 to 100,000 people whereas if we talk about the true information then it barely reaches 1000 people. With respect to this research problem, we also came to know that politicians and stock marketers use these types of practices to achieve their agenda, or we can say people generally use such methods to get their work done, make profits, or gain power.

A.Misinformation : The basic difference between misinformation and disinformation is the intent of the person or outlet sharing it. Misinformation includes incorrect or misleading content such as conspiracy theories, hoaxes, click-bait headlines, and fabricated reports. Its goal is to shape or alter public opinion on a given topic.

B. Disinformation: Fabricated reports, clickbait, hoaxes can spread the disinformation. The area of concern is that even educated individuals read news from any media source and forward it without verifying or looking for a valid source of information. The large amount of information available on social media, combined with the short attention period of readers, can allow fake information to go unchecked. Machine learning is empowering PCs to handle assignments that have, up to this point, just been completed by individuals. It is a domain in which PCs are given the ability to comprehend or learn just like humans do.

Neural System works like a human cerebrum. Neural System has various neurons interconnected with one another. The learning procedure of the neural system is like a human mind i.e. it learns by models. The neural system has numerous applications. The hidden pattern and information about an issue can be utilized to anticipate future circumstances or occasions and play out a wide range of complex dynamics.

In the current online social network, there are a great deal of issues such as fake profiles, online imitation, impersonation, and so forth. The current scenario has shown that no work has been done yet to provide an efficient way to tackle the challenge of fake news and fake profiles [22]. In this paper we aim to solve this problem by giving the system auto programmed identification of fake profiles and texts so that the social activity of individuals becomes more secure and by utilizing this technique, we can make it simpler for others to deal with fake news and fake accounts, which were not possible before physically. From a data mining perspective, the survey addresses relevant areas of study, open problems, and future directions of study. Research directions are shown in Figure 1.



2. RELATED WORK DONE

Different ML models have been trained with metadata by Wang et al [18]. The author primarily used convolutional neural networks (CNN). Shu et al. [12] explored veracity assessment to discover fake news online. Network analysis approach and linguistic cue approach are explored as assessment methods. Integrating these methods results in a stronger hybrid strategy for identifying fake news online. An approach discussed by Vosoughi et al. [19] focuses on spread of morphed news and analysed how its diffusion on Twitter differs from that of real news. The study by Ahmed et. [20] extracted linguistic features from text data and trained multiple machine learning models like support vector machine, decision tree, K-nearest neighbour, logistic regression where support vector machine and logistic regression achieves highest accuracy of around 92%. Kon taxis et al. (2011) depicts a model of the product that targets discovering whether the profile of a specific client was cloned from one online informal community into another by contrasting attributes of the profiles having comparable qualities among a few online interpersonal organizations. A Saberi et al. (2007) proposed gathering strategies to distinguish phishing tricks. Information mining arrangement calculations such as Naive Bayes, K-nearest neighbour, and Poisson probabilistic hypothesis and Naive Bayes are accustomed to ordering spam and non-spam. The combination of these two classifiers is used to achieve higher accuracy.

Naive Bayes, k-nearest neighbor, and Poisson datasets of authentic images to learn the distribution of genuine image features. [20].

2.1 GANs Effectiveness: GANs have shown great effectiveness in various domains, particularly in tasks involving data generation. Their ability to produce realistic and high-quality data has revolutionized several fields:

1. Image Generation: Image Generation: GANs are capable of producing highly realistic images. Applications include creating photorealistic faces, artwork, and even super resolution images.
2. Data Augmentation: GANs can generate additional training data, especially when the original dataset is small calculation independently give precision of 87%, 88.3.5%, and 91.2% individually. After teaming up these three methods, it gives a higher accuracy of 93.8%. The precision to recognize the tricks can be improved by utilizing different strategies, for example, Neural Network Systems and SVM. Yumen Qin et al [19] utilized the Naive Bayes classifier. Data Sources include Twitter, Facebook, and other social media platforms. The accuracy that they achieved was very low because the data on these sites were not 100% credible.

The Generative Adversarial Network (GAN) complements the fully unsupervised approach used in conjunction with the Autoencoder to generate high dimensional feature vectors from news sentences. GANs can be trained on large or imbalanced. This is particularly useful in medical imaging, where collecting large amounts of labelled data is challenging.

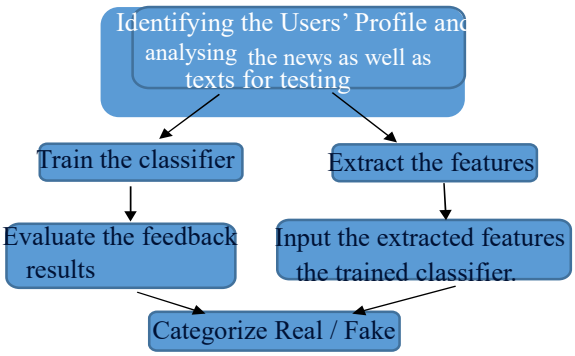
3. Anomaly Detection: GANs can learn the distribution of normal data, making them effective at detecting anomalies by recognizing samples that deviate from the learned distribution.

3. ADOPTED METHODOLOGY

The research method adopted to detect fake news and profiles is explained in the Figure 2 below.
In this step-wise process, firstly the identification of suspicious users' profile are selected. Then the features are extracted. Pass the extracted features into the trained classifier. The trained classifier would classify that into real or fake. The result and feedback act an input and the classifier will be trained again. The classifying techniques used are Random Forest, Neural Network, Support Vector Machine, LSTM and Naïve Bayes'.

3.1 Random Forest

As its name suggests, there are some trees based on the different subsets of the dataset. An average is calculated to enhance the prediction accuracy of the dataset. It is supervised learning which is utilized for classification. Instead of depending on a single tree, it takes decisions from each tree.
Ensembles use the divide-and-conquer strategy to improve performance and act as a form of nearest-neighbour predictor.



3.2 Support Vector Machine (SVM) SVM is an algorithm that classifies an isolating hyperplane. Ultimately, the calculation

provides an optimal hyperplane to classify the different models.
Hyperplane separates the plane for each class by diving into 2 regions in 2d space.

Support Vector Machine algorithm reasonably isolates these classes. Data points to the left of the line are the green circle, while data points to the right falls into the blue square. SVM does the detachment of classes.

3.3 NAÏVE BAYES

There is a micro chance in your life that you've never heard of this theorem. It turns out that this theorem finds its way into machine learning, becoming one of the highly decorated algorithms. Naive Bayes is a classification algorithm for binary and multiclass characterization issues. Rather than calculating the probabilities of each attribute, they are assumed to be conditionally independent given the class value. Overall, the methodology performs shockingly well on information where this suspicion does not hold.

3.4 Neural Network A neural network is what it says in the name. It is a cluster of neurons that are utilized to process data. They get information, process it, and likewise yield electric signs to the neurons it is associated with and utilize biomimicry. Long- term memory is a subset of the artificial architecture of neural networks that is used to process multiple data points in images, speech, audio, and text.

3.5 LSTM

Long term memory architecture processes image data points, text, speech, and audio. It consists of an input gate, a forgetting gate and a gate of output with one cell, as shown in figure 3. The vanishing gradient problem is also addressed using Recurrent Neural Networks (RNN) that are trained in supervised and unsupervised ways.

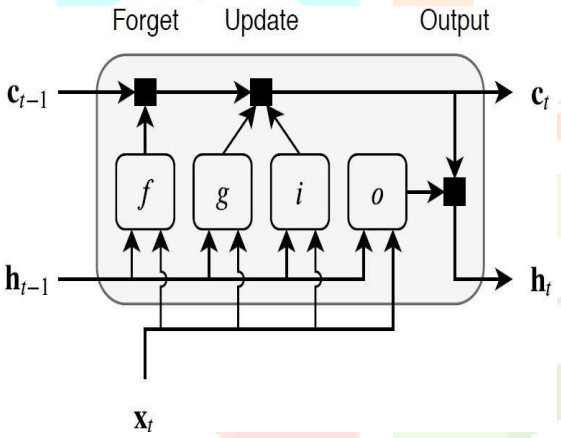


Fig 3: LSTM model (source: internet)

4. EXPERIMENT AND RESULT ANALYSIS

Implementation is sorting an object into a specific class based on the training dataset used to train the classifier. The classifier is trained on a dataset to identify similar objects with the highest precision and accuracy. A classifier is a kind of algorithm that is utilized for classification purposes. In this paper, we have utilized 3 classifiers, specifically NN, SVM, and RF, for the detection of fake profiles, and for the fake news, we have used LSTM and Naïve Bayes and have, in this manner, compared their efficiencies and accuracies.

Some of the modules/libraries implemented in the research are NumPy, Skit, and Pandas. For the IDE we have utilized Google Collab. It is a free opensource platform that is online hence no installation is required and has all the required libraries.

Step 1: Data Collection and pre-processing of data. Step 2: Generate false or fake profiles (accounts) and fake news.

Step 3: Validation of Data to discover fake and genuine profiles, also the data validation is done.

Step 4: New features are created according to the data set. Step 5: Apply neural networks,

random forest and SVM, LSTM, Naïve Bayes' to detect tampered profiles.
Step 6: Calculate precision (accuracy), review and recall parameters.

4.1 Data set

We have a need for a dataset of fake and real/genuine profiles. Different features as mentioned in table 1, used in the dataset are the number of followers, friends, and the count of their status. The Classification is used for training data set and efficiency of the algorithm is calculated by the testing of the data set. From the dataset utilized, more than 70 percent of profiles are utilized to train the data, and 30 percent of profiles to test the data.

TABLE 1: USED SET OF FEATURES FOR FAKE PROFILES

S.no	**Features**
1.	Number of friends
2.	Number of followers
3.	Preferred Count
4.	Sex code
5.	Listed Count
6.	Languages Known
7.	Status Count

TABLE 2: EXTRACTED FEATURES OF USER'S PROFILE

Attribute	**Explanation**
Post Count	Fake Accounts have a low count of the average no of posts.
Followers Count	Fake Accounts have low followers count or high follower counts of the same group.
Comment Count	Fake accounts share untrusted
	links and advertisements.
Events	Fake accounts do not share the event and live locations frequently.
Location	Fake accounts have irrelevant locations.

Tagged Post	Fake accounts have less number of tagged posts.
Created Time	Fake accounts use the timeline for a shorter period of time.
Description	The description is used to connect with more number of people.

Although online news can be collected from various sources, it is a challenging task to manually determine the variety of news. Because of those challenges, existing public data sets of fake news are rather limited.

(a) Frequent word in true article

(b) Frequent words in fake article

Dataset contains the news article's frame, the news article's title and an article's mark and subtitle. The datasets were used from the Kaggle and GitHub.

4.2 Confusion Matrix

It summarizes the prediction results of the classification problem, or it can be said that the performance of the classification algorithm can be summarized using this. This compares the different positives and negatives. This proposes the techniques wherein the classification model is confused while it makes predictions The figure 4 shows the normalized confusion matrix.

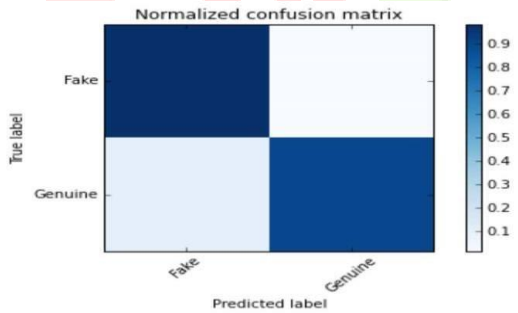
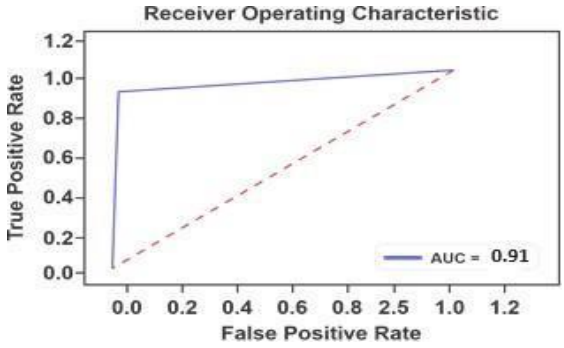


Fig 4: Normalized Confusion matrix of Neural

The mistakes performed by the classifier however extra significantly the variety of errors which can be done. Normalized implies that every one of these groupings is spoken to as having 1.00 examples. Therefore, the aggregate of each column in a fair and normalized matrix is 1.00, on the grounds that sum of each row speaks to 100% of the components in a specific subject, bunch, or class. Normalized Confusion Matrix is shown in the below table 3.

TABLE 3: NORMALIZED CONFUSION MATRIX



NAIVE BAYES RESULTS: For detection of fake news.

The results are shown using the confusion matrix. After performing the Naïve Bayes model on our dataset, an accuracy of 89% is achieved.

LSTM Results: Now we are moving on towards some discussion about the results that we obtained using LSTM for detection of fake news.

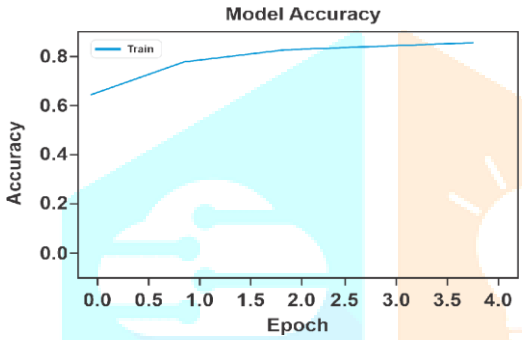


Fig 5: Accuracy of LSTM

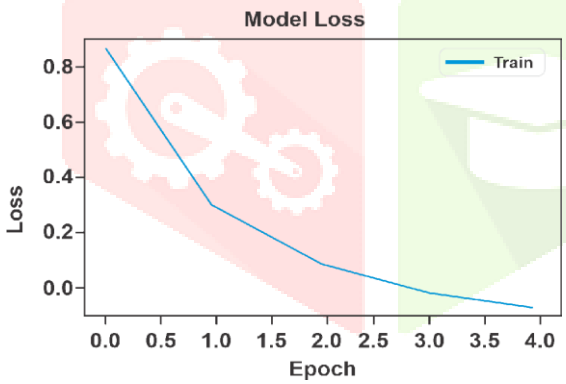


Fig 6: Loss of LSTM model

It shows that the accuracy of the model is increasing after every iteration shown in figure 5. The model is gradually learning, and the weights are being updated with the least loss percentage as shown in above figure 6.

Neural Network	0.98367347	0.016326531
	0.10377358	0.896226421
Random Forest	0.988880597	0.011194031
	0.10135135	0.898648651
SVM	0.97761194	0.02238806
	0.16216216	0.83783784

4.3 AUC – ROC

It presents performance estimation for classification problems at different threshold limits. ROC probability and AUC indicate the extent or degree of separation between different classes and represent how well the model is suited to differentiating between them. The curve is plotted between TPR and FPR as shown in figure 7 and figure 8.

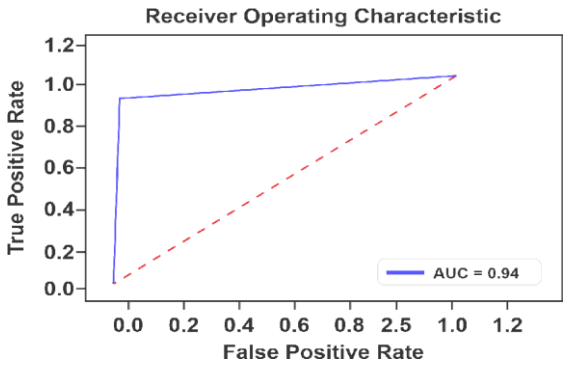


Fig 7: ROC curve Neural Network

5. RESULT ANALYSIS

For detection of fake profiles online, we utilized Kera's with TensorFlow backend using python to execute this model. The method which was implemented in our research has successfully and efficiently rectified the nature of profiles with the methodologies discussed in the above section. We have obtained graphs which show the value we have achieved during the testing part in our datasets. This value is nothing, but it validates the value having scalar nature which is the attempt we have made during the time for training of the dataset. Subsequently, it distinguishes if the profile is genuine or fake. The general accuracy over all of ML models was high with the most elevated being 94.3% utilizing Neural Networks and 94% utilizing Random Forest strategy lastly 90.01% utilizing SVM calculation algorithm. For detection of fake news Python language was the most used machine learning tool. All experiments are in python. Another method is programming. Fake news dataset includes four functions as ID, title, text, and label and having 7796 entries. Naïve bayes model shows an accuracy of 89%. As observed, the loss decreases with each epoch. After performing LSTM model, it shows an accuracy of 94% as shown in
Figure 9.

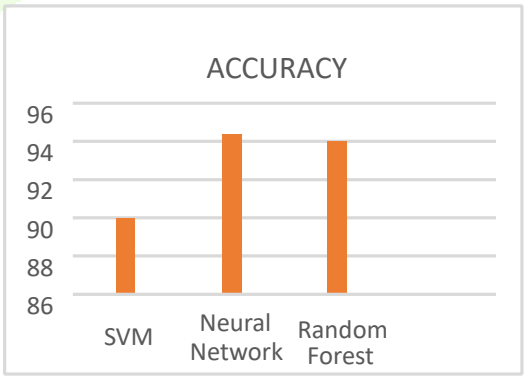


Fig 8: ROC curve SVM

Navie bayes and LSTM experiment conducted. After seeing the test, we found that naïve bayes show 89% accuracy while LSTM shows 94% accuracy with the dataset that we used. A newly emerging research area is detecting fake news on social media platforms. stats and explained how our algorithms works too, then showed the results of Naïve.

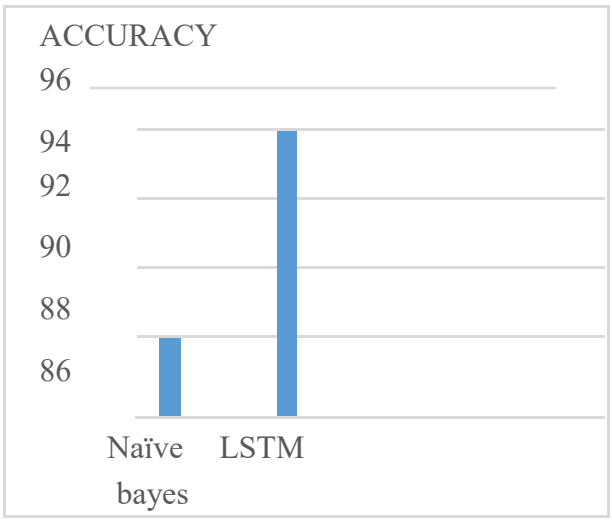


Fig 9: Comparative analysis between classifiers for fake news analysis

6. CONCLUSION AND FUTURE SCOPE

Fake accounts on social media exist for different reasons by people. The outcomes are about distinguishing whether the profile is fake or real by utilizing built highlights and trained using ML models for the detection of fake profiles. The prediction demonstrates that the algorithm neural networks system has an accuracy of 94.3%. Machine learning approach is proposed for detecting fake profiles, where our framework arranges a bunch of fake profiles to decide if they have been made by a similar entertainer. Our assessment of both in-test and out-of-test information indicates solid execution. Social media has become increasingly prevalent, large number of people consumes news from social media. It also disseminates fake news; however, it has a significant bad impact on users and the population. As discussed, the fake news is determined by analysing current literature in two phases: detesting and identifying. We have also discussed our dataset and its

REFERENCES

1. M. Saberi, M. Vahidi, and B. M. Bidgoli, “Learn to detect phishing scams using learning and ensemble methods,” in *IEEE*, 2007, pp. 311– 314.
2. D. K. Srivastava , L. Shambhu, “Data classification using support vector machine”, *J. Theor. Appl. Inf. Technol.* (JATIT), 2009.
   M. Alsaleh, A. Aarifa, A. M. AlSalman, M. Alferez, and A. Almutairi, "TSD: Detecting Sybil accounts in Twitter," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Detroit, MI, USA, 2014, pp. 463-469, Doi: 10.1109/ICMLA.2014.81.
3. Y. Shen, J. Yu, K. Dong, and K. Nan, “Chinese micro-blogging system,” in *Springer*, 2014, pp. 596–607.
4. M. S. B. Main, “Research paper on basic of artificial neural network,” *Int. J. Res. Inf. Technol. Compute. Commun.*, vol. 2, no. Jan., pp. 96–100, 2014.
5. M. Egale G. Stringline, and G. Vigna, “Towards detecting compromised accounts on social networks,” *IEEE*, vol. 5971, no. c, 2015.
6. B. Hudson, J. Matthews, S. Gura Jala, J. S. White, B. Hudson, and J. N. Matthews, “Fake Twitter accounts: Profile
   characteristics obtained using an
   activitybased pattern detection approach,”

ACM, no. Aug., 2015. 8. Y. Bosham and K. Benzos, “Thwarting fake OSN accounts by predicting their victims,” in *Proc. 8th ACM Workshop Arif. Intel. Secur.* (Ayse '15), 2015, pp. 81–89.
S. Rahman, T. Huang, H. V. Mahatha, and M.Fallouts, “Detecting malicious Facebook applications,” in *IEEE/ACM*, 2015, pp. 1 15.

S. Rahman, T. Huang, H. V. Mahatha, and M. Fallouts, “Detecting malicious Facebook applications,” in *IEEE/ACM*, 2015, pp. 1–15.

9. K. B. Kansara, “Security against Sybil attack in social networks,” in *Proc. ICICES*, 2016.

10. M. Malign, “Identity verification mechanism for detecting fake profiles in online social networks,” *Int. J. Commun. Newt. Inf. Secure.*, no. Jan., pp. 31–39, 2017.

11. L. Bilge, T. Strufe, D. Baluarte, and E. Karda, “All your contacts are belonging to us: Automated identity theft attacks on social networks,” in *Proc. 18th Int. Conf. World Wide Web*, Madrid, Spain, 2009, pp. 551– 560.

12. L. Jin, H. Takai, and J. B. Joshi, “ Towards active detection of identity clone attacks on online social media,” in *Proc. ACM Conf.*, San Antonio, TX, USA, Feb. 2011, p. 27.

13. M. Conti, R. Poovendran, and M. Secchi Ero, “Fakebook: Detecting fake profiles in online social networks,” in *Proc. 2012 IEEE/ACM Int.
Conf. Adv. Soc. Newt. Anal. Mining*
(ASONAM), Turkey, 2012, pp. 1071–
1078.

14. S. Gura Jala, J. S. White, B. Hudson, and J. N. Matthews, “Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection,” in *Proc. Int. Conf. Soc. Media Soc.* (Society '15), Toronto, Ontario, Canada, 2015.

15. W. Y. Wang, “Liar, liar pants on fire: A new benchmark dataset for fake news detection,” in *Proc. Assoc. Compute. Linguist.*, Stroudsburg, PA, USA, 2017.

16. S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146– 1151, 2018.

17. H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using ngram analysis and machine learning techniques,” in *Proc. Int. Conf. Intel. Secure Dependable Syst. Diatribe. Cloud Environ.*, Vancouver, Canada, 2017, pp. 127–138.

18. Y. Qin et al., “Predicting future rumours,” *Chin. J. Electron.*, vol. 27, no.
3, pp. 514–520, May 2018, Doi:
10.1049/cje.2018.03.008.

19. P. Bhardwaj, K. Yadav, H. Alsharif, and R. A. Abdalla, “GAN-based unsupervised learning approach to generate and detect fake news,” in *Proc. Int. Conf. Cyber Secure., Privacy, Netw.* (ICSPN 2022), Lecture
Notes Newt. Syst., vol. 599, Springer,
Cham, 2023, Doi: 10.1007/978-303122018-0_37.

20. M. D. Molina, S. S. Sundar, T. Le, and D. Lee, “‘Fake news’ is not simply false information: A concept explication and taxonomy of online content,” *Am. Behave. Sci.*, vol. 65, no. 2, pp. 180–212, 2021, Doi: 10.1177/0002764219878224.

21. E. Aimer, S. Amri, and G. Brassard, “Fake news, disinformation and misinformation in social media: A review,” *Soc. Newt. Anal. Mining*, vol. 13, no. 1, 2023, Doi: 10.1007/s13278-023-01028-5.