

Automated detection of fake news

Eslam Fayez, Amal Elsayed Aboutabl, Sarah N. Abdulkader

Department of Computer Science, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

Article Info

Article history:

Received Jul 26, 2022

Revised Aug 7, 2022

Accepted Sep 11, 2022

Keywords:

Content based detection systems

Fake news

Fake news detection systems

Fake news types

Hybrid based detection systems

ABSTRACT

During the last decade, the social media has been regarded as a rich dominant source of information and news. Its unsupervised nature leads to the emergence and spread of fake news. Fake news detection has gained a great importance posing many challenges to the research community. One of the main challenges is the detection accuracy which is highly affected by the chosen and extracted features and the used classification algorithm. In this paper, we propose a context-based solution that relies on account features and random forest classifier to detect fake news. It achieves the precision of 99.8%. The system accuracy has been compared to other commonly used classifiers such as decision tree classifier, Gaussian Naïve Bayes and neural network which give precision of 98.4%, 92.6%, and 62.7% respectively. The experiments' accuracy results show the possibility of distinguishing fake news and giving credibility scores for social media news with a relatively high performance.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Eslam Fayez

Department of Computers Science, Faculty of Computers and Artificial Intelligence, Helwan University

Ain Helwan, Helwan, Cairo, Egypt

Email: eslamfayez_csp@fci.helwan.edu.eg

1. INTRODUCTION

Nowadays, people spend a large portion of their time on the internet, mostly accessing websites and using social media for knowing the news. Fake news can spread quickly and significantly over websites and social media possibly causing social, economic, and political disturbances. Hence, the need to detect fake news among the huge amount of spreading news emerged. Fake news is written and used to look like real news to deceive the reader who normally does not check for the reliability of the sources or the arguments in the content of the news. The widespread of the Internet has led to the emergence of a new era and phase in the study of fake news. Social media not only provides a powerful environment for sharing information, but also allows data to be collected from large numbers of participants.

Fake news has been exploited in many fields as in politics. Fake news has affected the results of the 2016 US presidential elections [1]. It was reported that Saudi Arabia financed the presidential campaign of Emmanuel Macron [2]. Moreover, fake news has been used in promoting products or defamation for another product [3]. It is often created for commercial interests in order to attract viewers and generate advertising revenue. Due to the popularity of social media websites, the tendency to create spam accounts has increased. Therefore, dealing with accounts is particularly important to detect fake news. A number of features are used to detect spam accounts such as the number of followers and the number of followings.

Fake news and rumors take advantage of social media and communication technologies to expand. For example, Twitter, one of the most popular social media platforms, has around 313 million active users each month, which post each day around 500 million tweets [4]. This environment attracts the spammers and their attention, they use Twitter for other purposes they use Twitter for malicious purposes as spreading malware and phishing legitimate users. Spammers can use uniform resource locators (URLs) inside Twitter

to make advertisement, can follow/unfollow legitimate users aggressively. They may use trending topics to get the attention of the users. Dealing with fake news has become inevitable to limit and reduce its spread.

Fake news detection systems can be categorized into three categories: content-based systems, context-based systems, and hybrid systems (Figure 1). Content-based systems rely on the text to decide whether the information within the text is true or false. Linguistic features of the text are used including lexical and syntactic features. N-gram (a contiguous sequence of n items from a given sample of text or speech) and bag of words are used to obtain these features. In context-based systems, the focus is directed to the user or the source that published the news. Here, user features include followers and following, profile username, posts or tweets, and many others. A hybrid system is a mix of both content-based and context-based systems relying on both text and source.

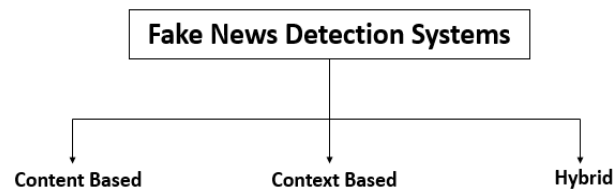


Figure 1. Fake news detection systems

In the case of detecting fake news, according to the content can find many works in this area. Chopra *et al.* [5] work towards the headline and the body stance detection and detect the type of the relation between the headline and the body. First used support vector machine (SVM) to detect whether the headline-article pairing is related or unrelated. If related can use bidirectional conditionally encoded long short-term memory (LSTMs) with bidirectional global attention to getting the type of the relation (agree, disagree, and discuss), achieving 86.58% accuracy. Rakholia and Bhargava's [6] worked on the same problem using LSTM based recurrent neural network (RNN) model achieving 88.38%. Thorne *et al.* [7] worked in the area of getting the stance relation between the headline and the body using five independent classifiers as two layer classifier achieving 97.25% accuracy. Riedel *et al.* [8] targeted the area of stance detection getting the relation between the headline and the body achieving 81.72% accuracy. Research by Bajaj [9] can predict news is fake or real based only on its content. Using multiple different model implementations, using convolutional neural network (CNN) with max pooling and attention achieving 97%. Research by Volkova *et al.* [10] built linguistically-infused neural network models to classify social media posts from accounts. Classifying posts into verified and suspicious categories: hoax, propaganda, click-bait and, satire achieving 95%. Rashkin *et al.* [11] trained an LSTM model that takes the sequence of words as the input and predicts whether it's mostly true or mostly false achieving 65%. Ahmed *et al.* [12] used a solution to detect fake news by using a machine learning ensemble approach. They use different textual properties not normal properties to detect fake content. By using different machine learning algorithms and make a combination of them with various ensemble methods getting results up to 99%.

In the context based detection fake news can be detected using the account that published the news using some account features (number of hashtags, number of URLs, and number of mentions) [1], [13]–[15]. There are many works in this area Tacchini *et al.* [16] said that they can classify posts of Facebook as real or hoaxes based on the user who liked the post using two approaches classification via logistic regression and classification via harmonic boolean label crowdsourcing achieving above 99% accuracy, but having big problem case when no likes to the post. Kabakus and Kara [17] detected spam on social media using features as account features (user name, profile photo, and number of tweets), post (tweet) features (mentions, hashtags), and graph features relations between users and their posts (tweets) can be represented as graphs achieving above 99% accuracy. According to Gee and Teh [18] the first step was identifying an initial set of features that could be used by the learning algorithm to distinguish between spammer profiles and normal user profiles as (followers-to-following ratio, following-to-followers ratio), then training data collection, they collected data using Twitter application programming interface (API) and then tried to implement a Naïve Bayes learning algorithm then implemented a linear classifier SVM achieving 89.6%. Research by Lin and Huang [14], they only use two features the URL rate and the interaction rate after finding that some features are not so effective in the detection. URL rate, they find that the ratio for the normal users and spammers is spaced, where normal users rate only 7% for the spammers is 95% in the collected tweets. Interaction rate is an effective feature because accounts of normal users usually interact with friends while spam accounts do not have this interaction only post URL links. Using these two features achieving accuracy reach 88.5%.

In the case of hybrid based fake news detection, which make a combination of the previous two categories (Figure 2). Conroy *et al.* [4] proposed the idea of using a hybrid model, claiming that it could improve the accuracy they talk about content based detection and network (context) based detection. Ruchansky *et al.* [19] develop a model that combines content and context, building capture, score, integrate (CSI) model. Capture which interacts with text and Score which interacts with users and combine them in the Integrate which demonstrate the quality of CSI on two real world datasets Twitter and Weibo datasets in the case of Twitter achieving 89.2% and in the case of Weibo achieving 95.3%. Benevenuto *et al.* [20] targeting content and user attributes achieving 87.6%. Yang *et al.* [21] made an analysis of the tactics used by Twitter spammers for evasion, and then design several new and robust features to detect Twitter spammers. The tactics used by the spammers to evade existing detection approaches is two types: profile-based feature and content-based feature. They achieve 87.7%.

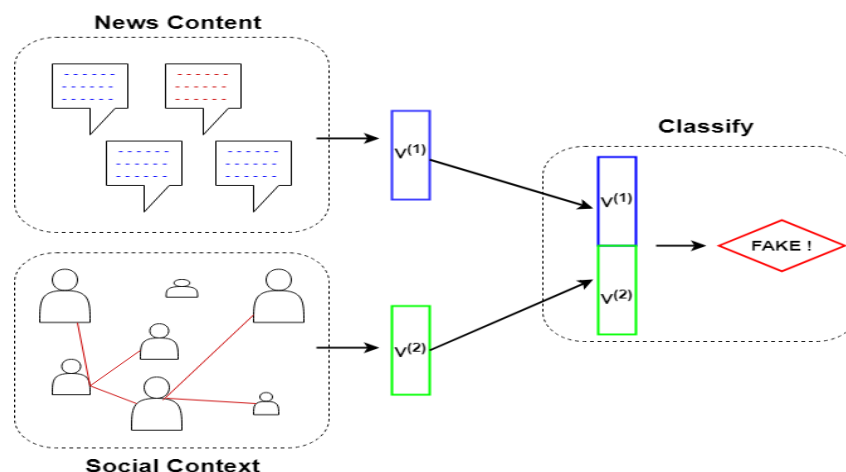


Figure 2. Hybrid model

Based on the fact that the best results were conducted via context based solutions [1], [13]–[15], our proposed solution uses the source account details with random forest classifier in order to determine news credibility. Account features can be used to indicate whether the news is true or false. Account features that are used as the number of hashtags, number of URLs, number of mentions, number of followers, friends and count to achieve best results. Our approach depends on specific account features not normal account features, our approach uses the most significant features that can enhance the detection.

- URL count feature for each user is highly distinctive as stated in [14] which denotes that the more URLs in use, the higher the probability of account fakeness. (Positive relation)
- Hashtag count is one of the influencing features, which means the number of hashtags the account uses in the posts. (Positive relation)
- Followers count is another important feature, which is expected to be less compared to real accounts [17] because no one wants to be a follower to a fake account. (Inverse relation)
- Friends count which means friends that account has, real accounts can have friends greater than fake accounts. (Inverse relation)
- Mention count which is the number of mentions use in the posts.
- Retweet count is the number of retweets for the post.

2. METHOD

The proposed solution consists of five main stages (Figure 3), the first stage is data preprocessing which prepares raw data for further processing, data preprocessing is a main component, used for preparing the dataset and encoding it in a form that the algorithm can parse. Data preprocessing takes care of null and missing values. Dataset normalization is an important step in data preprocessing, normalization goal is to get a common scale from original values and not deforming the difference between the ranges of values. Train test split is very important in the data preprocessing phase because the model before being deployed must be evaluated.

The second stage is feature extraction which get a lower-dimensional space from the existing features, which means original features are transferred to lower features, to enhance the classifier efficiency

must find a set of features that is most informative and compacted. Moreover, to fulfill reliable classification must extract features from the main ones. The third stage is feature selection also called attribute selection or variable selection. It is the process of automatically select the most relevant attributes in the dataset related to our work predictive modeling problem. To reduce modeling computational cost and enhance model performance it is recommended to make the number of input values minimum as much as possible. The fourth stage is the classifiers which specify the classifiers that will be used on the dataset which will be explained in the experiment subsection.

The fifth stage is the prediction stage that there is a model that is able to detect fake accounts. Our model starts with preparing data for the next stages. In preparing data our model takes care of null values and normalizes data to a common scale. Then applying feature extraction technique principal component analysis (PCA) which means original features are transferred to lower features, to enhance the classifier efficiency must find a set of features that is most informative and compacted. Then using the feature selection technique (SelectKBest) to reduce modeling computational cost and enhance model performance it is recommended to make the number of input values minimum as much as possible. Then using the classifier getting best results by using random forest classifier getting results up to 99% compared with other classifiers used as neural network which get results up to 77%, Gaussian Naïve Bayes classifier get results up to 81%, and decision tree classifier which get up to 98%.

Our dataset is a collection of Twitter accounts, real and fake accounts collected by [22]. Dataset collected from different sources, sources for real accounts and sources for fake accounts. The source for real accounts is #elezioni2013 this dataset consists of 1,481 real accounts (which means physical human). The sources for fake accounts were 1,000 from <http://twittertechnology.com>, 1,000 from <http://fastfollowerz.com>, and 1,000 from <http://intertwitter.com>.

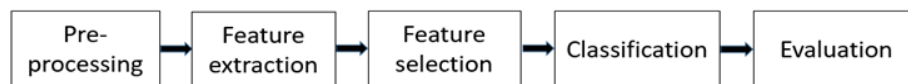


Figure 3. Fake news detection system

3. RESULTS AND DISCUSSION

Using random forest classifier in our experiment with the account features to reach the best results, and then comparing random forest results with commonly used classifiers as SVM, decision tree, Gaussian Naïve Bayes, and neural network. Before using classifiers, there are some phases data must pass through it as data preprocessing, feature extraction, and selection. Random forest classifier has been used classify the collected Twitter accounts. Using the total dataset with total of 1,481 real accounts and 1,884 fake accounts. Getting the best results 99%.

Then comparing random forest results with others classifiers find that neural network classifier with the same dataset getting results reaching 77%. After that Gaussian Naïve Bayes classifier is used to get accuracy to reach 81%. Now its decision tree turns which gets good results to reach up to 98.2%. SVM classifiers were used but the results were not any good. All classifiers are used with the same data and same environment, summarizing the results (Table 1). Classification experiments are performed using train test split with test size equal to 0.20 indicates the percentage of the data that should be held over for testing. It's usually around 80/20 or 70/30.

Table 1. Accuracy results

	Precision	Recall	F1 score
Random forest	.997	.984	.99
Neural network	.627	.997	.77
Gaussian Naïve Bayes	.926	.72	.81
Decision tree	.984	.981	.982

3.1. Classification results

As seen in Table 2 the confusion matrix which shows the results of our experiment using random forest. The numbers presented in the table are relative to the total number of Twitter accounts. The results show that 378 of real accounts classified as real accounts and 202 of fake accounts are classified as fake accounts.

Table 2. Random forest classification results

		Predicated	
		Real	Fake
Real	Real	378	6
	Fake	1	202

Comparing our results with others find that our results are much better and reaching high and high accuracy for detecting accounts. Comparing our results with Azab *et al.* [15] using the same dataset with different designs find that they are using random forest reaching up to 96.1% and our result using random forest reaching up to 99.7%. Lee *et al.* [13] results reach 99.2%, Lin and Huang [14] getting up to 88.5%, and many others works, so if comparing our result with others results can find that our results much better.

4. CONCLUSION

According to the statistics which prove that getting information to become from online websites and social media. This makes the topic of fake news gain more attention because fake news can be published through online websites and social media easier and quicker than journals and other types of news publishing tools. According to the effects of fake news in society. It became necessary to deal with fake and rumor news. Trying to detect and know whether this news is rumor or real ones.





REFERENCES

- [1] A. K. Chaudhry, D. Baker, and P. Thun-Hohenstein, "Stance detection for the fake news challenge: Identifying textual relationships with deep neural nets," *CS224n: Natural Language Processing with Deep Learning*, 2017, pp. 1-117.
- [2] V. Singh, R. Dasgupta, D. Sonagra, K. Raman, and I. Ghosh, "Automated fake news detection using linguistic analysis and machine learning," in *Proceedings of the 2017 International Conference on Social Computing, Behavioural-Cultural Modeling, & Prediction and Behaviour Representation in Modeling and Stimulation*, 2017, pp. 1-3.
- [3] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, Springer Verlag, 2017, pp. 127-138, doi: 10.1007/978-3-319-69155-8_9.
- [4] N. K. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1-4, Jan. 2015, doi: 10.1002/pa2.2015.145052010082.
- [5] S. Chopra, S. Jain, and J. M. Sholar, "Towards automatic identification of fake news: Headline-article stance detection with LSTM attention models," *Proc. Stanford CS224d Deep Learn. NLP Final Project*, 2017.
- [6] N. Rakholia and S. Bhargava, "Is it true?-Deep learning for stance detection in news," Technical Report. Stanford University, California, USA, 2016.
- [7] J. Thorne, M. Chen, G. Myrianthous, J. Pu, X. Wang, and A. Vlachos, "Fake news stance detection using stacked ensemble of classifiers," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, 2017, pp. 80-83, doi: 10.18653/v1/W17-4214.
- [8] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, "A simple but tough-to-beat baseline for the fake news challenge stance detection task," Jul. 2017, [Online]. Available: <http://arxiv.org/abs/1707.03264>
- [9] S. Bajaj, "The pope has a new baby ! fake news detection using deep learning," *Fake news detection using deep learning*, 2017.
- [10] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2017, vol. 2, pp. 647-653, doi: 10.18653/v1/P17-2102.
- [11] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2931-2937, doi: 10.18653/v1/D17-1317.
- [12] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, pp. 1-11, Oct. 2020, doi: 10.1155/2020/8885861.
- [13] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots + machine learning," in *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*, 2010, pp. 435-442, doi: 10.1145/1835449.1835522.
- [14] P. C. Lin and P. M. Huang, "A study of effective features for detecting long-surviving Twitter spam accounts," in *2013 15th International Conference on Advanced Communications Technology (ICACT)*, 2013, pp. 841-846.
- [15] A. El Azab, A. M. Idrees, M. A. Mahmoud, and H. Hefny, "Fake accounts detection in twitter based on minimum weighted feature," in *ICDAR 2016: 18th International Conference on Document Analysis and Recognition*, 2016, pp. 13-18, doi: 10.5281/zenodo.1110582.
- [16] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," *arXiv preprint arXiv:1704.07506*, Apr. 2017, doi: 10.48550/arXiv.1704.07506.
- [17] A. T. Kabakus and R. Kara, "A survey of spam detection methods on Twitter," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 3, pp. 29-38, 2017.
- [18] G. Gee and H. Teh, "Twitter spammer profile detection," Available online: cs229.stanford.edu/proj2010/GeeTeh-TwitterSpammerProfileDetection.pdf, 2010.
- [19] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Nov. 2017, vol. Part F1318, pp. 797-806, doi: 10.1145/3132847.3132877.
- [20] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in *7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, CEAS 2010*, 2010, pp. 1-10.





- [21] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers," in *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 8, 2011, pp. 318–337, doi: 10.1007/978-3-642-23644-0_17.
- [22] "The Fake Follower Classifier Project," 2019. [Online]. Available: <http://wafi.iit.cnr.it/theFakeProject>. (accessed: Sep. 5, 2022).

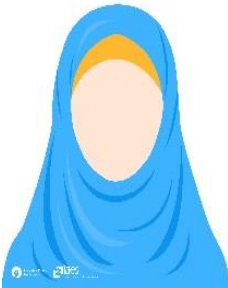
BIOGRAPHIES OF AUTHORS







Eslam Fayeze     received the B.Sc. Computer Science degree from Helwan University, Egypt, in 2013 and now preparing for the M.S. degree. His research interests include fake news detection, machine learning, and pattern recognition. He can be contacted at email: eslamfayeze_csp@fci.helwan.edu.eg.



Amal Elsayed Aboutabl     is currently a professor of Computer Science and the vice dean for community service and environmental development at the Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt. She received her B.Sc. in Computer Science from the American University in Cairo and both of her M.Sc. and Ph.D. in Computer Science from Cairo University. She worked for IBM and ICL in Egypt for seven years. She was also a Fulbright Scholar at the Department of Computer Science, University of Virginia, USA. She can be contacted at email: amal.aboutabl@fci.helwan.edu.eg.



Sarah N. Abdulkader     an Assistant Professor at Helwan University, graduated in 2005, obtained Master's degree in 2011 for research in the field of Vehicular Ad-hoc Networks, became a member of HCI Lab in 2013. She obtained PhD degree in 2015. Her PhD research work focuses on the exploration and incorporating of brain signals in security systems. It uses the non-invasively collected brain waves for claimed-identity authentication. Her current research interests include HCI, pattern recognition, and machine learning. She can be contacted at email: Nabil.sarah@gmail.com.