

A SMART SYSTEM FOR FAKE NEWS DETECTION USING MACHINE LEARNING

Anjali Jain¹, Avinash Shakya², Harsh Khatter³, Amit Kumar Gupta⁴
^{1,4}KIET Group of Institutions ^{2,3}ABES Engineering College, Ghaziabad

Abstract: Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites. It is harmful for the society to believe on the rumors and pretend to be a news. The need of an hour is to stop the rumors especially in the developing countries like India, and focus on the correct, authenticated news articles. This paper demonstrates a model and the methodology for fake news detection. With the help of Machine learning and natural language processing, author tried to aggregate the news and later determine whether the news is real or fake using Support Vector Machine. The results of the proposed model is compared with existing models. The proposed model is working well and defining the correctness of results upto 93.6% of accuracy.

Keywords: Artificial Intelligence, Fuzzy Logic, Fuzzy Inference, Machine Learning, Naive Based Classifier, News, Prediction, Recommendation, Support Vector Machine (SVM).

I. INTRODUCTION

In Today's world, anybody can post the content over the internet. Unfortunately, counterfeit news gathers a lot of consideration over the web, particularly via web-based networking media. Individuals get misdirected and don't reconsider before flowing such mis-educational pieces to the most distant part of the arrangement. Such type of activities are not good for the society where some rumors or vague news evaporates the negative thought among the people or specific category of people^[1]. As fast the technology is moving, on the same pace the preventive measures are required to deal with such activities. Broad communications assuming a gigantic job in impacting the general public and as it is normal, a few people attempt to exploit it. There are numerous sites which give false data. They deliberately attempt to bring out purposeful publicity, deceptions and falsehood under the pretense of being true news. Their basic role is to control the data that can cause open to have confidence in it. There are loads of case of such sites everywhere throughout the world. Therefore, counterfeit news influences the brains of the individuals. As indicated by study Scientist accept that numerous man-made brainpower calculations can help in uncovering the bogus news.

Fake news detection is made to stop the rumors that are being spread through the various platforms whether it be social media or messaging platforms, this is done to stop spreading fake news which leads to activities like mob lynching, this has been a great reason motivating us to work on this project. We have been continuously seeing various news of mob lynching that leads to the murder of an individual; fake news detection works on the objective of detecting this fake news and stopping activities like this thereby protecting the society from these unwanted acts of violence.^{[1][3][5]}

The main objective is to detect the fake news, which is a classic text classification problem with a straight forward proposition. It is needed to build a model that can differentiate between "Real" news and "Fake" news. This leads to consequences in social networking sites like Facebook, Instagram, microblogging sites like Twitter and instant messaging applications like WhatsApp, Hike where these fake news gets a major boost and gets viral among people, around the country and globe.^[2] The proposed system helps to find the authenticity of the news. If the news is not real, then the user is suggested with the relevant news article.

II. RELATED WORK

2.1 Overview

There have been quite a several initiatives taken to achieve fake news detection:

- In 2018 three students of Vivekananda Education Society's Institute of Technology, Mumbai published their research paper on fake news detection. They wrote in their research paper, social media age has started in 20th century. Eventually the web usage is increasing, the posts are increasing, the number of articles are increasing. They used various techniques and tool to detect fake news like NLP techniques, machine learning, and artificial intelligence.^{[5][6][7]}
- Facebook and WhatsApp are also working on fake news detection as they wrote in an article. They have been working for almost one year, and it is currently under the alpha phase.^[2]
- Nguyen Vo student of Ho Chi Minh City University of Technology (HCMUT) Cambodia did his research on fake news detection and implemented in 2017. He used Bi-directional GRU with Attention mechanism in his project fake news detection; Yang et al. originally proposed this mechanism. He also used some Deep learning algorithms and

tried to implement other deep learning models such that Auto-Encoders, GAN, CNN.

- Samir Bajaj of Stanford University published a research paper on fake news detection. He detects fake news with the help of NLP perspective and implements some other deep learning algorithm. He took an authentic data set from Signal Media News dataset.

Several approaches have been taken to detect the fake news after massive widespread fake news in recent times. There are three types of fake news contributors: social bots, trolls, and cyborg users ^{[3][4]}. Social Bots says, if a social media account is being controlled by a computer algorithm, then it is referred to as a social bot. The social bot can automatically generate content. Secondly, the trolls are real humans who “aim to disrupt online communities” in hopes of provoking social media users into an emotional response. Other one is, Cyborg. Cyborg users are the combination of “automated activities with human input.” Humans build accounts and use programs to perform activities in social media. For false information detection, there are two categories: Linguistic Cue and Network Analysis approaches. The methods generally used to do such type of works are Naïve Bayes Classifier and Support Vector Machines (SVM).

2.2 Facebook Works to Stop Misinformation and False News

Facebook in an article quoted they are working to fight the spread of false news in two key areas. First is disrupting economic incentives because of most false news in financially motivated. Second one is, Building new products to curb the spread of false news ^[6]. Some of the preventive measures taken by facebook are mentioned here:

- **Ranking Improvements:** News Feed ranks reduce the prevalence of false news content.
- **Easier Reporting:** Determine what is valuable and what is not. Stories that are flagged as false by our community than might show up lower in the user feed

2.3 WhatsApp Work for Fake News Detection

To stop the spread of misinformation, WhatsApp has implemented some security measures and also fake news detection, though these are under alpha phase and are yet to be rolled out to the beta users. WhatsApp testing ‘Suspicious Link Detection’ feature: This feature will alert uses by putting a red label on links that it knows to lead to a fake or alternative website/news. Additionally, if a message has been forwarded from a device more than 25 times, the message could be blocked. ^[2]

2.4 Outcome

As mentioned in the above section, all top most giants are trying to hide their selves from the rumors and focus should be on true news and authenticated articles. More or less, the approaches follow in the extraction are based on machine learning and Natural language processing. The classifiers, models and analytical algorithms are required to work hand in hand for the authentication of news articles. SVM will be used

in the paper by the authors as an existing best suitable approach with Naïve Bayes. SVM is best suited for binary classification. There are various news websites and news blogs which allows to work with RSS feeds and import the references of the news articles. This will helps us in finding the news accuracy.

III. PROPOSED MODEL

There are three main modules in the proposed model, named as

3.1 Aggregator

News aggregator sites enable clients to see news and updates from different sources at one advantageous area. They get the information, arrange them in labels/classes, and show it organized appropriately for simpler utilization. There are various top most sites like google news, feedly, news360, etc. which provide the semi-structured news data. The RSS Aggregator plugins are available for make the things easy. Basically, aggregator builds the quality and accuracy of/in the news. Data gathering is the main purpose of any of the news aggregator. A basic methodology is consistently checking RSS Feeds; Article extraction from various news site and gather information. Regularly applied systems to discover related articles are watchword based methodologies. At the point when all procedure is done, they show important or ongoing news on the page. ^[15]

3.2 News Authenticator

New authenticator follows some steps to check whether the news is true or false. It will compare news which is given by our side with different websites and various news sources if that news is found on any news website then it shows the given news is true, else it shows there has been no such news in last few days. This can help us from fake news. These days’ fake news spread very fast because of social media and the internet. So, news authenticator helps us to detect either the given news is fake or real.

3.3 News Suggestion /Recommendation System

News suggestion suggests recent news and suggests the news related to the news which the user has given for authentication. If the news is fake, then this news suggestion gives the related news on that topic. The news suggestion suggests the news based on keywords which you give in your news which you wish to authenticate. ^[4]

IV. METHODOLOGY

4.1 Approach

Because of the multi-dimensional nature of fake news, the recognizing the category of news is not so easy. It is obvious that a practical technique must contain a few perspectives to precisely handle the issue. This is the reason the proposed strategy is a mix of Naïve Bayes classifier, Support Vector

Machines, and semantic investigation. The proposed strategy is completely made out of Artificial Intelligence draws near, which is basic to precisely order between the genuine or the fake, rather than utilizing calculations that can't mirror subjective capacities. The three-section strategy is a blend between Machine Learning calculations that subdivide into managed learning procedures, and characteristic language preparing techniques.

4.1.1 Naive Bayes

A Naive Bayes classifier is a supervised machine learning algorithm that uses Bayes' theorem. The variables that are used to generate the model are independent of each other. It is proven that this classifier itself provides pretty good results.^{[9][10]}

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i)$$

The classification is conducted by deriving the maximum posterior, which is the maximal $P(C_i|X)$ with the above assumption applying to Bayes theorem. This assumption greatly reduces the computational cost by only counting the class distribution. Naive Bayes is popular algorithm which is used to find the accuracy of the news whether its real or fake using multinomial Naïve Bayes. There are number of algorithms that focus on common principle, so it is not the only algorithm for training such classifiers. To check if the news is fake or real naïve Bayes can be used.

4.1.2 Support Vector Machine (SVM)

SVM is a good algorithm to extract the binary class based on the data given to the model. In the proposed model, the work is to classify the article in two categories either true or false. A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both regression and

classification purposes. It is based on the idea of finding the hyper-plane that best divides the dataset into two classes. Hyper-planes are decision boundaries that help the machine learning model classify the data or data points. How the classification of the data point is done using a hyper-planes can be seen in figure 4.2 depicted below:^{[11][12]}

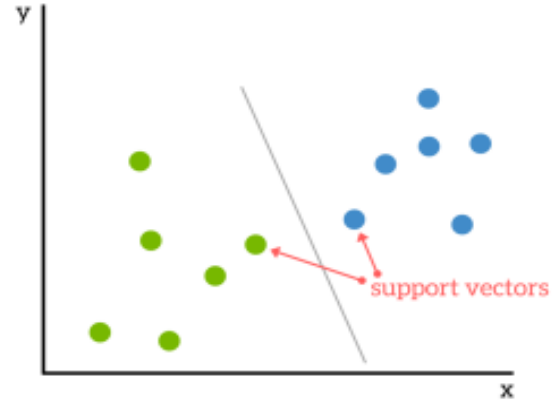


Figure 4.2 Depiction of hyper-plane dividing the dataset into two classes

Also, the benefits of utilizing the SVM strategy are that it will in general be exceptionally precise and performs incredibly well on datasets that are semi-structures structured. Moreover, this method is truly adaptable since it tends to be utilized to arrange or even decide numbers. Likewise, support vector machines have the capacity to deal with high dimensional spaces and will in general be memory proficient.^[18]

4.2 System Architecture

SVM is a good algorithm to extract the binary class based on the data given to the model. In the proposed model, the work is to classify the article in two categories either true or false. A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both regression and

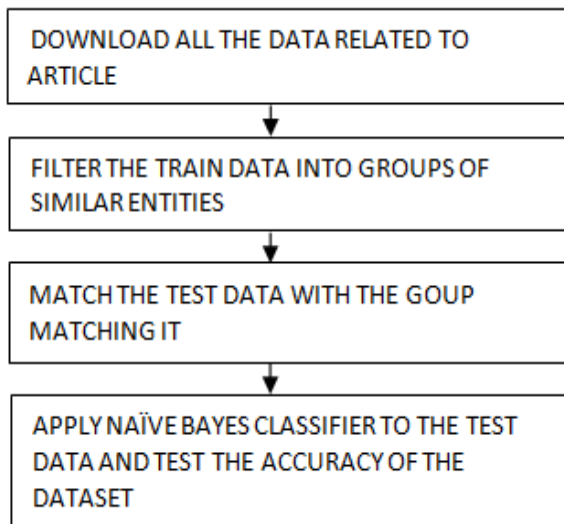


Fig 4.1 Use of Naïve Bayes Classifier

4.2 System Architecture

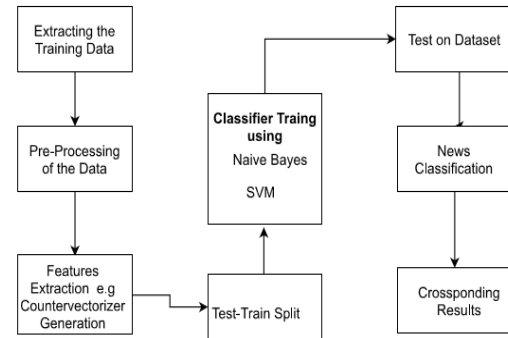


Figure 4.3 Flow chart – Classifier Training

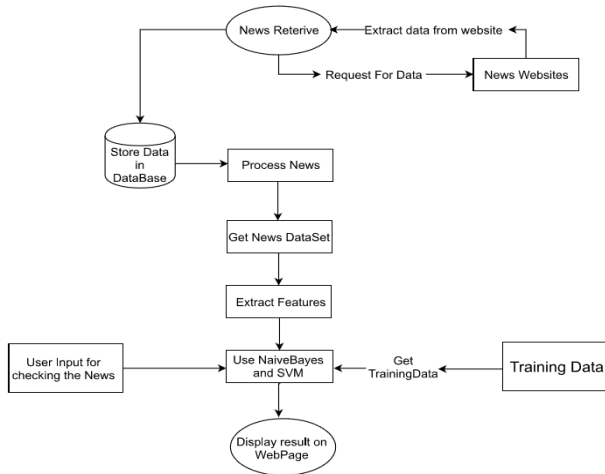


Figure 4.4 Flow chart – Proposed Model

V. IMPLEMENTATION AND RESULTS

For the implementation purpose, the four existing approaches are considered. The results of mentioned four models are compared with the proposed model, it is found the accuracy among top 200 results is mentioned in the table 5.1. The demonstration is done using python programming on R studio and some machine learning algorithm.

Table 5.1 Result Comparison

Article	Accuracy	Implementation Method
R. V. L, C. Yimin, and C. N. J (2016)	76%	NLP
M. Granik and V. Mesyura (2017)	74%	Naive Bayes
Y. Seo, D. Seo, and C. S. Jeong (2018)	86.65%	CNN
Jain A., Khatter H., Shakya A. (2019)	93.50%	Naive Bayes, SVM, NLP

VI. CONCLUSION

It is significant to find the accuracy of news which is available on internet. In the paper, the components for recognizing Fake news are discussed. A mindfulness that not all, the fake news will propagate via web-based networking media. Currently, to test out the proposed method of Naïve Bayes classifier, SVM, and NLP are used. In future, ensuing algorithm may provide better results with hybrid approaches for the same purpose fulfilment. The mentioned system detects the fake news on the based on the models applied. Also it had provided some

suggested news on that topic which is very useful for any user. In the future, the efficiency and accuracy of the prototype can be enhanced to a certain level, and also enhance the user interface of the proposed model.

REFERENCES

- [1] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE 1st Ukr. Conf. Electr. Comput. Eng. UKRCON 2017 - Proc., pp. 900–903, 2017.
- [2] <https://indianexpress.com/article/technology/social/whatsapp-fight-against-fake-news-top-features-to-curb-spread-of-misinformation-5256782/>
- [3] A. Martínez-Garcia, S. Morris, M. Tscholl, F. Tracy, and P. Carmichael, "Case-based learning, pedagogical innovation, and semantic web technologies," IEEE Trans. Learn. Technol., vol. 5, no. 2, pp. 104–116, 2012.
- [4] P. R. Humanante-Ramos, F. J. Garcia-Penalvo, and M. A. Conde-Gonzalez, "PLEs in Mobile Contexts: New Ways to Personalize Learning," Rev. Iberoam. Tecnol. del Aprendiz., vol. 11, no. 4, pp. 220–226, 2016.
- [5] T. Granskogen and J. A. Gulla, "Fake news detection: Network data from social media used to predict fakes," CEUR Workshop Proc., vol. 2041, no. 1, pp. 59–66, 2017.
- [6] R. V. L, C. Yimin, and C. N. J, "Deception detection for news: Three types of fakes," Proc. Assoc. Inf. Sci. Technol., vol. 52, no. 1, pp. 1–4, 2016.
- [7] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News," pp. 7–17, 2016.
- [8] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel Visual and Statistical Image Features for Microblogs News Verification," IEEE Trans. Multimed., vol. 19, no. 3, pp. 598–608, 2017.
- [9] S. Gilda, "Evaluating machine learning algorithms for fake news detection," IEEE Student Conf. Res. Dev. Inspiring Technol. Humanit. SCORED 2017 - Proc., vol. 2018–January, pp. 110–115, 2018.
- [10] Y. Seo, D. Seo, and C. S. Jeong, "FaNDeR: Fake News Detection Model Using Media Reliability," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, vol. 2018–October, no. October, pp. 1834–1838, 2019.
- [11] S. Das Bhattacharjee, A. Talukder, and B. V. Balantrapu, "Active learning based news veracity detection with feature weighting and deep-shallow fusion," Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017, vol. 2018–January, pp. 556–565, 2018.
- [12] S. Helmstetter and H. Paulheim, "Weakly supervised learning for fake news detection on Twitter," Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2018, pp. 274–277, 2018.
- [13] S. B. Parikh, V. Patil, and P. K. Atrey, "On the Origin, Proliferation and Tone of Fake News," Proc. - 2nd Int. Conf. Multimed. Inf. Process. Retrieval, MIPR 2019, pp. 135–140, 2019.
- [14] A. Dey, R. Z. Rafi, S. Hasan Parash, S. K. Arko, and A. Chakrabarty, "Fake news pattern recognition using linguistic analysis," 2018 Jt. 7th Int. Conf. Informatics, Electron. Vis. 2nd Int. Conf. Imaging, Vis. Pattern Recognition, ICIEV-IVPR 2018, pp. 305–309, 2019.
- [15] N. Kim, D. Seo, and C. S. Jeong, "FAMOUS: Fake News Detection Model Based on Unified Key Sentence Information," Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, vol. 2018–November, pp. 617–620, 2019.
- [16] R. L. Vander Wal, V. Bryg, and M. D. Hays, "X-Ray Photoelectron Spectroscopy (XPS) Applied to Soot & What It Can Do for You," Notes, pp. 1–35, 2006.
- [17] M. Gahirwal, "Fake News Detection," International Journal of Advance Research, Ideas and Innovations in Technology, vol. 4, no. 1, pp. 817–819, 2018.