IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11368

Fake news detection using machine learning

Mrs. Namrata S. Khade, Abhishek.N.Nikhade, Rani Samrit, Madhulika Bhanarkar, Kalpesh Kathale

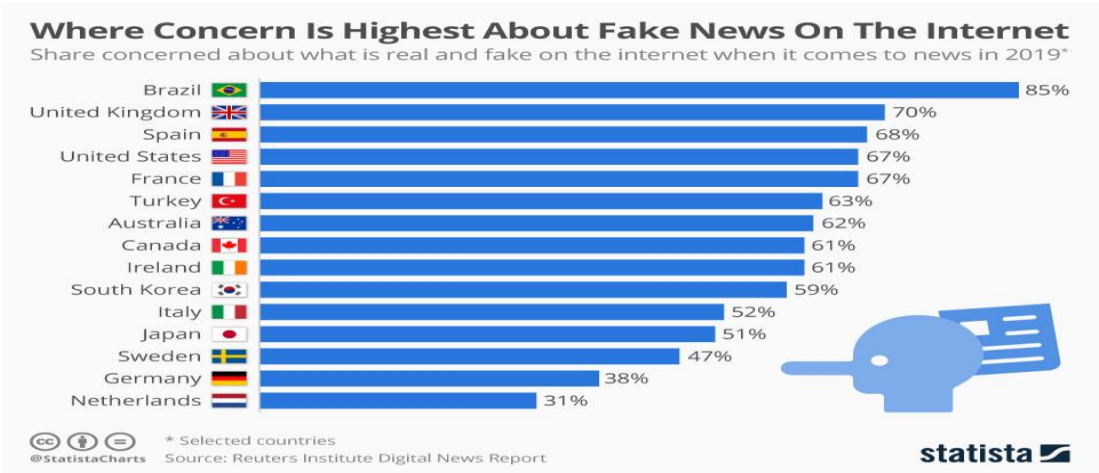
Dept. of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India

Abstract: Ever since the internet was introduced fake news had also grab their hand around it, if we think carefully fake news being around since the old times but now only internet has expanded up the process.in the old times it took days for news to travel across lands but with the help of internet it happens within seconds we can share anything instantly. Social media also plays an important role in this it is a wide spreading and is a matter of serious concern due to ability to cause a lot of social and national damage. The credibility of social media networks is also at stack where the spreading of fake information is prevalent. Our paper review is tackling this by using machine learning and artificial intelligence to detect if the news is fake or not, we used more than 5 algorithms to get maximum accuracy towards our results, algorithms we used are decision tree, random forest, support vector machine (svm), naive bayes and KNN (k-nearest neighbours) which then show the credibility of news is later shown in percentage

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

I. INTRODUCTION

The rise of fake news has become a global problem now a days that can even major tech companies like twitter and google are struggling to solve. Fake news can influence people's perceptions. It can be difficult to determine whether a text is true or false. In Today's world, anybody can post anything they want over the internet. Individuals get misdirected and don't reconsider before flowing such mis-educational pieces to the most distant part of the arrangement. The prevalence of fake news has increased recently with the rise of social media, especially the Facebook new feeds, and this misinformation has gradually seeped into the mainstream media. Several factors have been implicated in the spread of fake news, such as political polarization, post truth politics, motivated reasoning, conformation bias, and social media algorithm. Fortunately, there are number of computational techniques that can be used to mark certain articles as fake on the basis of their textual content. The news media evolved from newspapers, and magazines to a digital form such as online news platforms, blogs, social media and other digital media formats. It is easier for consumers to acquire the latest news on their fingers. Fake news detection is made of two parts authenticity and intent. Authenticity means that fake news contains false information that can be verified as such, which means that theory is not included in fake news as they are difficult to be proven true or false in most cases. The second part, intent, means that the false information has been written with the goal of misleading with reader.



This data shows us the percentage where concern is highest about fake news on the internet.

Data like this show us the reality of news weather we are consuming a true news or not, finding weather a news is true or not will take a lot of time and research not everyone got that much time in this busy world. That is one of the main reasons of fake news expansion people just look at a certain news and then share it no one confirms the news because no

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11368

one got that much time, this is also a negligence of mob. Fake news has also caused many riots in our country the problem of fake news can have dangerous consequences and its impact was witnessed during the lockdown when a panic situation was created through social media.

II. PROPOSED SYSTEM:

The methods we used in our model - tf idf is the basis for the pa algorithm. It denotes to term frequency and inverse document frequency. TFIDF may be used to separate stop-words in a variety of subject-NN ALGORITHM The K-NN supervised machine learning algorithm denotes to K nearest neighbour algorithm. This algorithm is better in the case of a small data set and pa is better for the large dataset. KNN can be used for classification and regression problems also it significantly slows when the amount of data increases.PA ALGORITHM-Pa algorithm denotes to passive aggressive algorithm. Passive-aggressive algorithms are generally used for large scale learning also it is one of the few online learning algorithms. A very good example of this algorithm would be to detect a fake news on a social media website like twitter, where new data is being added every second. So, the purpose of our project is to developed a method for detecting the fake news stories using natural language processing. First, we gathered our data, pre-processed the text, and then translated our articles into supervised model features. Our goal is to develop a model that classifies a given news article as either true or fake. Eliminations is that our system does not guarantee 100% accuracy. Basically, our model determines the status of inputted news and give us the result weather the news is fake or not with the accurate percentage which may not be 100% right but our model offers 70% accuracy throughout.

III. RESEARCH METHOD

Fake news has a multi-dimensional nature the recognizing the category of news is not so easy. That is why we used proposed strategy mix with Naïve Bayes classifier, Support Vector Machines, and semantic investigation. Our strategy is completely based on machine learning it is basic to precisely order between the genuine or the fake, skipping utilizing calculations that can't mirror subjective capacities. This three-section strategy is a mix between Machine Learning calculations that subdivide into managed learning procedures, and its characteristic with language preparing techniques.

I. Naive Bayes

A naïve bayes classifier completely supervised with machine learning algorithm that use bayes theorem. The variable is independent of each other that we used in our model. Hats why classifier give us good results.

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

This classification is conducted by deriving the maximum posterior with the maximal P (Ci |X) with the above bayes theorem. Our assumption greatly reduces the computational cost which basically do is count the class distribution. Naïve bayes is one of the most popular algorithms which is used to find the accuracy of the news whether its real or fake using multinomial Naïve Bayes. There are many algorithms that focus on common principle, naïve bayes is not the only algorithm to test whether a news is fake or not and also not the only algorithm to test such type of classifier.

II. Support Vector Machine (SVM)

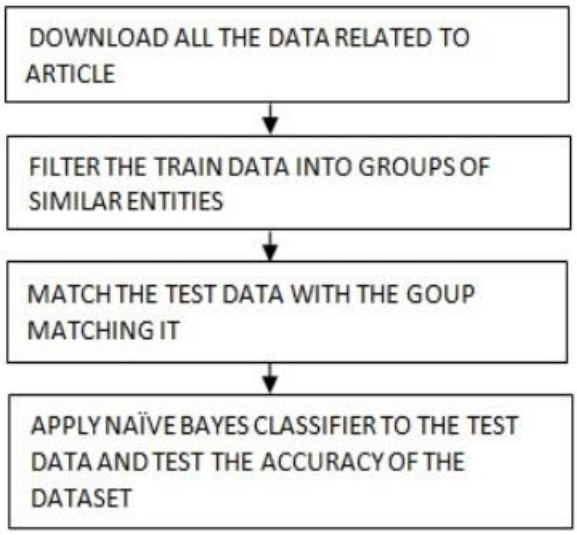
SVM is a very good algorithm if u want to extract the binary class based on the given data model. In our proposed model, the work is classified into two categories either true of fake. A Support Vector Machine (SVM) is a supervised machine learning algorithm which is used for both regression and classification purposes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11368



The idea is based on finding the hyper-plane that divides the datasets into two classes. Hyper-planes are nothing but decision boundaries that help the machine learning to classify the data or data points. The classification of data points on hyper-plane is shown in the fig:

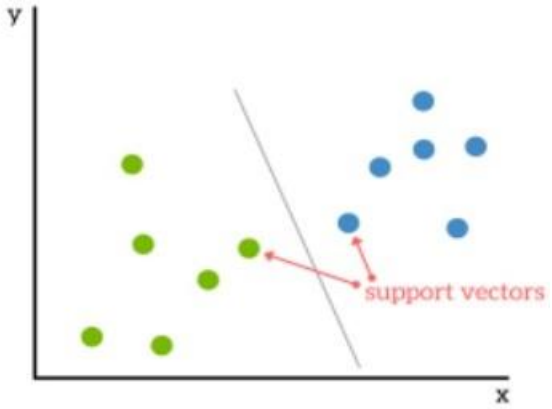


Fig: Depiction of hyper-plane dividing the dataset into two classes

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.

System Architecture

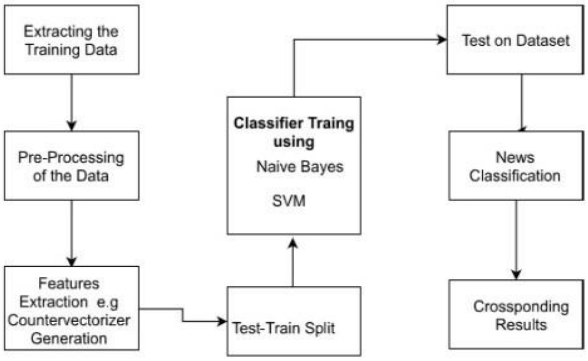


Fig: Flow chart – Classifier Training

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11368

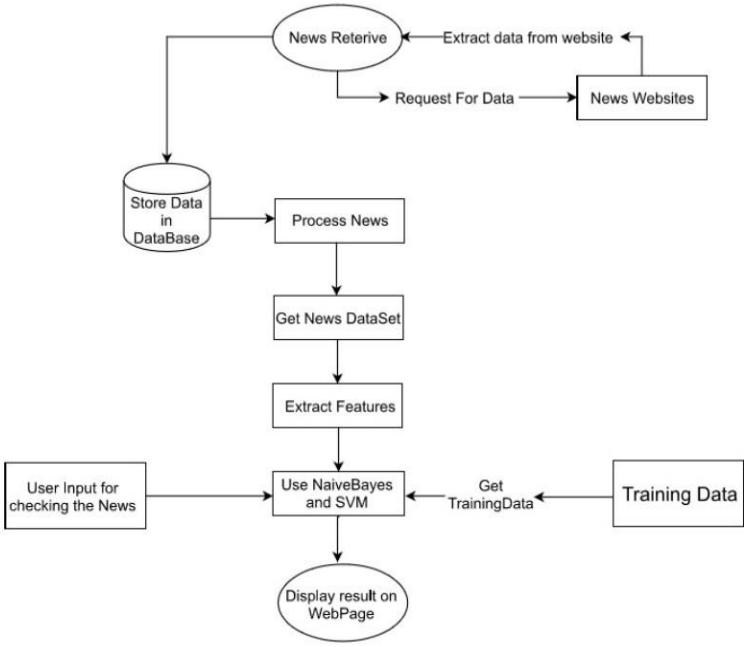


Fig: Flow chart – Proposed Model

IV: IMPLEMENTATION AND RESULTS

For our implementation propose, the five methods we used are considered. The results of the five algorithms are mentioned below.

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**

It is found that the accuracy in our dataset is maximum when combined. The demonstration is done on jupyter notebook python.

V: CONCLUSION

So, in conclusion we came that when we use a df-idf vectorizer and have implemented the fake news detection model using k n and pa classifier pa algorithm accuracy increases when the training data increases and k n algorithm accuracy decreases. The accuracy of p classifier is higher so we selected this as the best machine learning model to detect the fake news so overall the performance for our data set was better with pa classifier.

This research focuses on detecting the fake news by reviewing it in two stages: Disclosure and characterization. In first stage, the basic content and principles of fake news are highlighted in social media. During the discovery stage, the current methods are reviewed for detection of fake news using different supervised learning algorithms.

Our research focus on betterment for people now a day in this busy life no one confirm the news they get people just look at a certain news and they forward it without thinking of consequences weather it might be fake and it can cause a major riot. Many of the major incident are cause because of fake news to minimize this we tried to make a certain program for it which can incredibly help the people.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11368

REFERENCE

[1]. Economic and Social Research Council. Using social media. Available at: https://esrc.ukri.org/research/impact-toolkit/social-media/using-social-media

[2]. Gil, P. Available at: https://www.lifewire.com/what-exactly-is-twitter-2483331. 2019, April 22.

[3]. HTTPS://TRAILHEAD.SALESFORCE.COM/EN/CONTENT/LEARN/MOD ULES/VIRTUAL-WHITEBOARDING-QUICK-LOOK/LEARN-HOW-TOWHITEBOARD-DURING-A-VIRTUAL-MEETING ʊ

[4]. HTTP://OPENCV-PYTHONTUTROALS.READTHEDOCS.IO/EN/LATEST/PY_TUTORIALS/PY_GUI/PY_VIDEO_DISPLAY/PY_VIDEO_DISPLAY.HTML▪

HTTPS://IEEEXPLORE.IEEE.ORG/ABSTRACT/DOCUMENT/5989909

[5]. Researchgate.net. Available at: Available at: https://www.researchgate.net/figure/Pseudocode-ofnaive-bayes-algorithm_fig2_325937073. 2018. [

[6]. Researchgate.net. Available at: https://www.researchgate.net/figure/Pseudocode-for-KNNclassification_fig7_260397165, 2014.

[7]. Rampersad G, Althiyabi T 2020 “Fake news: Acceptance by demographics and culture on social media” J. Inf. Technol. Politics 2020, 17, 1–11.

[8]. Naphaporn Sirikulviriya; Sukree Sinthupinyo. “Integration of Rules from a Random Forest.” International Conference on Information and Electronics Engineering (p. 194: 198). Singapore: semanticscholar.org. 2011.

[9]. Jasmin Kevricet el. “An effective combining classifier approach using tree algorithms for network intrusion detection.” Neural Computing and Applications, 1051–1058. 2017.

[10]. Shivam B.Parikh and Pradeep K.Atrey. “Media-RichFake News Detection: A Survey.” IEEE Conference on Multimedia Information. Miami, FL: IEEE. 2018.

[11]. Mykhailo Granik and Volodymyr Mesyura. “Fake news detection using naive Bayes classifier.” First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Ukraine: IEEE. 2017.

BIOGRAPHY

Mrs. Namrata S. Khade is working as a Asst. Professor at Priyadarshini College of Engineering. She is having 12 years of experience in the field of teaching to engineering students. She completed her Engineering in 2007 and Master in Engineering in 2013.She is a member of IEEE, ISTE and CSI. She is having more than 30 research published in International Journals and Conferences. Her interests include distributed parallel computation, System Programming, Computer Graphics and Wireless Sensor Network.