Detection of Fake news using Machine Learning Algorithms

Axit Kachhia[1], Prof. Bhargesh Patel[2], Dr. Miral Patel[3]

[1,]Dept- ME - Information Technology, G H Patel College of Engineering and Technology, CVM University, Anand–388001, Gujarat, India
[2.] Assistant Professor, Dept- Information Technology, G H Patel College of Engineering and Technology, CVM University, Anand–388001, Gujarat, India
[3.] Associate Professor, Dept- Information Technology, G H Patel College of Engineering and Technology, CVM University, Anand–388001, Gujarat, India

---****************---

ABSTRACT

In the world of social media and smartphone uses, people are tended to believe what that see on the social media. It is observed that many fake and unauthenticated videos and content is viral within fraction of seconds and people are believing on that news also. To deal with such type of the situation we developed one model in which the news can be analyzed and then it can be categorized as an authenticate news and fake. To do so machine learning algorithms are used for checking the accuracy of the model.

Keywords—Fake news, Stop words, Natural Language Processing, Machine Learning, Classification

INTRODUCTION

We know that now a days there is a tremendous use of internet, and there are many things people are getting to know from the sources of internet. It is required to know that each and every source or a data coming from the internet should be valid and cannot be fake, so that people who are used to get information from those internet sources are able to get the authenticate data. Here in this project we aimed to find the data, specifically news from the social networking and we are trying to analyze that data to check whether it's been authenticated or a fake news.

Here we connected with the social network AIP to get the content from it and then we train the data using supervised algorithm analyze the data and used the confusion matrix to validate the data. Misleading information is being forwarded by fake news that needs to be checked. Although the scope is limited because it also depends on the manual checking by the human [5]. Fake news can make negative impact on a society, individuals (Generally young age people) can be distracted via fake news, so it is important to check and authenticate the news before moving ahead and forwarding that particular news. Finding out the truth and fake news can make the positive impact on the society.

In order to find the reality in words, we need to check the occurrence of the words in sentence. People are expressing their emotions via the post they are providing on the social network [1]. When the news or any information has been altered than linguistic approach is used to analyze that data. Linguistic Cue approaches notice fake news by contain going the information exploiters in the writing style of the news content.

In Such cases labeling the data using supervised learning algorithms are useful, but one of the disadvantages of the supervised learning algorithm is, it requires the significant amount of the labeling on the data [2]. If we are doing the identification of a fake news in a real time than it would-be time-consuming task for us to labeled that real time data. Fake news can be exploiting by sharing the post, re-twitting the actual content with tempering the data.
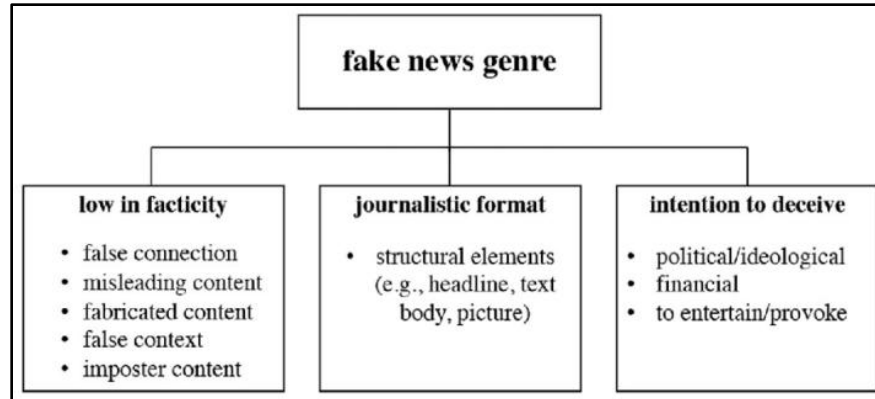
Classification of Fake news

Visual Based Fake News: This type of news contains media and graphical representatives to showcase the data and those type of data is also containing the information in the form of image file or in the form of graphs and charts [4]. It is more popular among the users on social networks like Facebook, Twitter and Instagram to share these types of news. Recently we have got the feature of forwarded in WhatsApp messenger in order to prevent the spread of fake news.

Linguistics based Fake News: This type of the fake news is based on the text communications between or group of the users [4]. If we are taking about the e-mails, posts, website and the blogs than the contain post by the user is unsupervised and easily can be received as wrong information [4].

User based Fake News: This news is based on the fake accounts that have been created on the social network and it is mainly targeting to some age group with specific gender [3]. This news is generally targeting the same topic and spreading the news on social media.

NATURAL LANGUAGE PROCESSING

The fundamental rationale for using NLP is to examine one or more system or algorithm specialties. An algorithmic system's Natural Language Processing rating enables the integration of voice understanding and speech production. It might also be used to recognize actions in a variety of languages. Presented a new perfect system for extracting actions from English, Italian, and Dutch talks by combining multiple language pipelines such as Emotion Analyzer and Detection, Named Entity Recognition, Parts of Speech Taggers, Chunking, and Semantic Role Labeling, which made NLP a strong search subject.

Computational linguistics—rule-based human language modeling—is combined with statistical, machine learning, and deep learning models in NLP. These technologies, when used together, allow computers to process human language in the form of text or speech data and 'understand' it's full meaning, including the speakers or writer's intent and sentiment.

Sentiment Analysis: it is used to extract the information from the sentence. Specifically, it is used to extract the specific term and emotions from the sentence and dataset. Sentiment models database and meaning glossary for constructive and destructive words, with classifications ranging from -5 to 5. Parts of speech taggers tools for languages such as European languages are being researched in order to create parts of language taggers tools in languages such as Sanskrit, Hindi, and Arabic.

METHODOLOGY

There are various approaches to deal with the text-based data and identifying the patterns from it. Some of the approaches that can be deal with the fake news identification are as follows:

a) **Language based approach:** This approach emphases on the use of linguistics by a human or software platform to detect fake news. Most of the people accountable for the spread of fake news have control over what their story is about, but they can frequently be uncovered through the style of their language [9]. There are many methods which are dealing with the language-based approach which are as follows:

b) **Bag of Words (BOW):** In this approach, each word in a paragraph is considered of equal importance and as independent entities. The frequency of individual terms is examined in order to detect symptoms of misinformation. N-grams are another name for these formats.

c) **Semantic Analysis**: Its truthfulness can be established by comparing personal experiences (such as restaurant ratings) with subject profiles derived from similar articles. Honest writers are more likely to make similar comments on the subject than other honest writers. This approach uses various compatibility assessments.

d)**Deep Syntax:** Deep syntax methods are performed through context-free stochastic grammars, and context-free stochastic grammars perform deep syntax tasks through analysis trees that allow context-free grammar analysis. Stochastic context-free grammar is an extension of context-free grammar.

e)**Topic-Agnostic Approach**: This category of approach recognizes fake news by considering topic-agnostic features rather than article content. This approach uses language and web markup features to identify fake news. Examples of off-topic features are 1) numerous ads, 2) long headlines with flashy phrases, and 3) evoking emotional reactions with text patterns that differ from mainstream news.

f)**Machine Learning Approach:** Fake news can be identified using machine learning techniques. This is accomplished by refining the algorithms using various types of training datasets. Computer scientists can use datasets to create new machine learning algorithms and techniques [11]. To train the algorithms to detect bogus news, datasets are used. What methods are used to create these datasets? One method is to use crowdsourcing [11]. Twitter has come up with a potential method for detecting and preventing the spread of false information via bogus accounts, likes, and comments [11]

g)**Knowledge based Approach:** Recent research suggests that to detect fake news, machine learning and knowledge engineering should be combined [12]. The speed with which fake news spreads on social media is a difficult problem with some of these fact-checking tools. Microblogging networks like Twitter allow for the rapid transmission of erroneous information to a large number of individuals [12].

h)**Expert oriented Fact Checking:** Expert-oriented fact checking requires careful analysis and checking of data and documentation [13]. Expert-based fact checking requires experts to manually assess the authenticity of news through investigations into specific claims and other investigations. Fact checking provides certainty about a particular item by comparing the accuracy of the text with other previously fact checked text [13].

i)**Crowd Oriented Fact Checking:** Crowdsourcing allows groups of people to make collective decisions by examining the authenticity of a message. The accuracy of the news is entirely based on the wisdom of the crowd. Kiskkit is an example of a platform that can be used for crowdsourcing. This platform allows groups of people to rate parts of a news article. After an article is evaluated, the crowd moves on to the next article for evaluation until the entire news article is evaluated and its accuracy is determined by the wisdom of the crowd.

j) **Hybrid Approach:** There are 3 commonly agreed upon factors of fake information articles, the primary detail is the textual content of an article, 2nd detail is the reaction that the articles acquired and ultimately the supply used that inspire the information article [10]. A current takes a look at has been carried out that proposes a hybrid version which enables to perceive faux information on social media via the use of an aggregate of human and gadget gaining knowledge of to assist perceive faux information.

PROPOSED SYSTEM

Using a social media now a days are not at all a bigger deal; everyone is used to spend the time on the social media and they are believing on the things they see on social media. The problem is people are not validating the content that is being provided on the social media. They are not concerned about whether the data or news is fake or not on social network. The main goal is to spread on a set of classification processes to obtain a classification model in order to be used as a scanner for a fake news by details of news discovery and embed the model in python to be used as a detection for the fake news data.
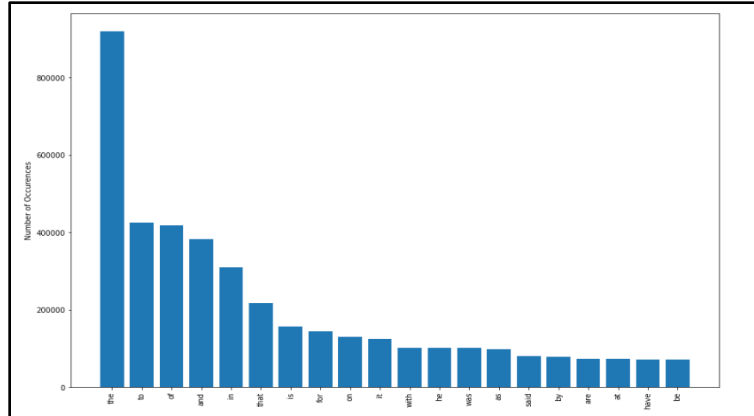
No. of Occurrence in stop words
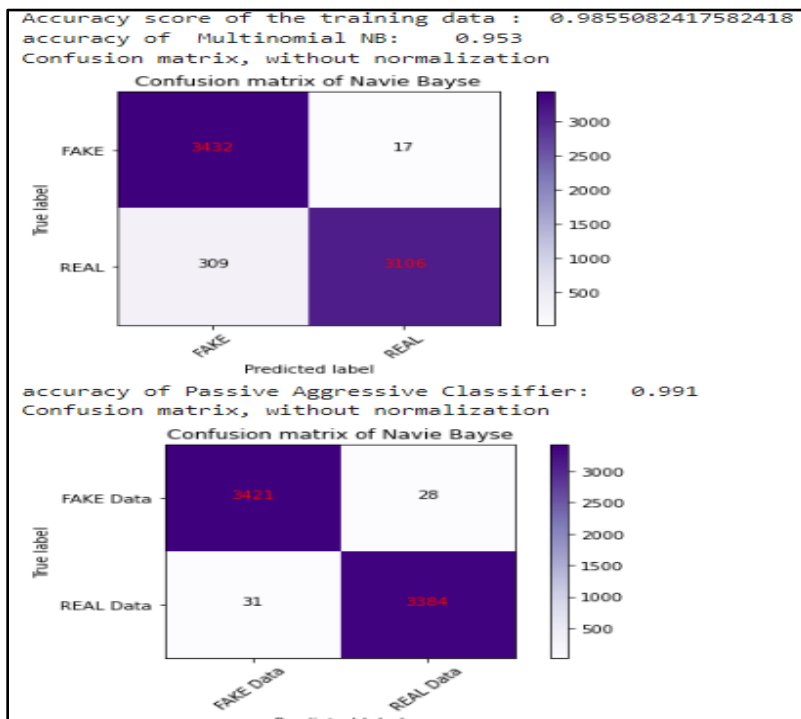
Multinomial NB Confusion Matrix

Final Accuracy score of the Algorithms for finding fake news.

CONCLUSION

In recent years, misleading content has been more prevalent, and its impact on online users has gotten worse. We developed a special frame to predict bogus news on social media in this research. The process of selecting features from datasets is critical because the ML algorithm relies on them. We have applied different algorithms to check the accuracy and prediction rate for finding the fake news. It is help to select the best model to find the accuracy.

FUTURE WORK

As we know that now a days data is the new oil, and the content of the internet is updating daily, We will try to apply the model on real time data set (yet it is a time-consuming task for a processing real time data) and we will try to get the results on the social networks like twitter and try to plot the real time graph with real and fake news. Although the big task is to store that Real time data and process that real time data, to do so the work can be expanded on the Hadoop cluster with the real time data processing. Also, the work can be expanded to authenticate the Image news also by using the image processing techniques.

REFERENCES

[1].   Miguel A. Alonso, David Vilares, Carlos Gómez-Rodríguez, JesúsVilares "Sentiment Analysis for Fake News Detection", MDPI Electronics, 2021

[2].   Siva Charan Reddy Gangireddy, Deepak P, Cheng Long, Tanmoy Chakraborty "Unsupervised Fake News Detection: A Graph-based Approach", ACM 2020.

[3].   Syed IshfaqManzoor, Dr Jimmy Singla, Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review" ICOEI, IEEE-2019

[4].   https://en.wikipedia.org/wiki/Detecting_fake_news_online

[5].   Z Khanam, B N Alwasel, H Siraf, M Rashid in "Fake News Detection Using Machine Learning Approaches, ASCI – 2020.

[6].   Uma Sharma, Sidarth Saran, Shankar M. Patil, "Fake News Detection using Machine Learning Algorithms" IJERT – 2021

[7].   https://libguides.tru.ca/fakenews/characteristics

[8].   Jana Laura Egelhofer, Sophie Lecheler, "Fake news as a two-dimensional phenomenon: a framework and research agenda", AICA – 2019.

[9].   Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Xiaoming Zhang, Zhoujun Li, Philip S. Yu, "TI-CNN: Convolutional Neural Networks for Fake News Detection" arXiv -2018

[10].   Ruchansky, N., Seo, S., Liu, Y.: CSI: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, pp. 797–806 (2017). 10.1145/3132847.3132877

[11].   Sivasangari V, Anand PV, Santhya R. A modern approach to identify the fake news using machine learning. Int. J. Pure Appl. Math. 2018;118(20):10

[12].   Qazvinian, V., Rosengren, E., Radev, D.R., Mei, Q.: Rumor has it: identifying misinformation in microblogs. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, pp. 1589–1599 (2011)

[13].   Ahmed, S., Hinkelmann, K., Corradini, F.: Combining machine learning with knowledge engineering to detect fake news in social networks - a survey. In: Proceedings of the AAAI 2019 Spring Symposium

[14].   Alkhodair S A, Ding S H.H, Fung B C M, Liu J 2020 "Detecting breaking news rumors of emerging topics in social media" Inf. Process. Manag. 2020, 57, 102018

[15].   Jeonghee Yi et al. "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques."In Data Mining, 2003. ICDM 2003. Third IEEE International Conference (pp. 427-434).

[16].   Ranjan et al. "Part of speech tagging and local word grouping techniques for natural language parsing in Hindi". In Proceedings of the 1st International Conference on Natural Language Processing (ICON 2003).

[17].   MonaDiab et al. Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks. Proceedings of HLT-NAACL 2004: Short Papers (pp. 149–152). Boston, Massachusetts, USA: Association for Computational Linguistics. 2004

[18].   Rouse, M. https://searchenterpriseai.techtarget.com/definition/machine-learning-ML May 2018

[19].   Adrien Benamira, Benjamin Devillers, Etienne Lesot, Ayush K. Ray, Manal Saadi, Fragkiskos D. Malliaros, "Semi-Supervised Learning and Graph Neural Networks for Fake News Detection", 2019 ACM/IEEE.

[20].   Gowri Ramachandran, Daniel Nemeth, David Neville, DimitriiZhelezov, BhaskarKrishnamachari,

WhistleBlower: Towards A Decentralized and Open Platform for Spotting Fake News", IEEE 2020.

[21]. Ankit Kesarwani, Sudakar Singh Chauhan, Anil Ramachandran Nair, "Fake News Detection on Social Media using K-Nearest Neighbor Classifier", IEEE – 2020.

[22]. Fake News Detection Using Deep Recurrent Neural Networks, Tao Jiang, Jian Ping Li, Amin UlHaq, AbdusSaboor, ICCWAMTIP, IEEE – 2020

[23]. Fake News Detection: A comparison between available Deep Learning techniques in vector space, Lovedeep Singh, IEEE- 2021.

[24]. Fake News Detection with Generated Comments for News Articles, Yuta Yanagi, RyoheiOrihara, Yuichi Sei, YasuyukiTahara, Akihiko Ohsuga, IEEE 2020.