FAKE NEWS DETECTION

ANGEL SHIBU[1], MINLA K S[2]

MSc Scholar, Computer Science, St Joseph's College (Autonomous) Irinjalakuda, Thrissur, India[1]

Assistant Professor, Computer Science, St Joseph's College (Autonomous) Irinjalakuda, Thrissur, India[2]

Abstract: Now a day's Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access. Because of its wide range of "fake news", been seen i.e., low quality news with intensity of false information. By the spread of this fake news, it has the potential for extremely negative effects on individuals and society. Therefor the reality of news publishing in the social media became more unfair and un relative. Detection a fake news in social media provide unique challenges that make existing detection algorithms from traditional news in media is ineffective or not applicable. First, fake news mainly written to mislead readers to accept the false information, which makes it difficult to detect based on news content. Second, one is that exploiting its auxiliary information is making more challenging in users by that social engagements with fake news make the data much big, incomplete, unstructured, and noisy.

Keywords: Detecting fake news, Alexa ranking, Google page ranking, protocol, comparison, content matching.

I.INTRODUCTION

Nowadays people tend to seek out and make out news from social media rather than traditional news organizations. The quality and important of news on social media is much lower than that of traditional news organizations. Big amount of fake news, i.e., news with intentionally false information, are produced online for a variety of purposes, such as financial and political gain. Fake news can have different type of effects on individuals and the society.

- First one is, people may be misled by fake news and accept the false information's.
- Secondly, fake news could change the way people respond to real news.
- Third, large number of fake news could break the reality of entire news ecosystem.

Thus, it is very important to detect the reality of the news that been published on social media. Fake news is intentionally made to mislead consumers, which makes it nontrivial to detect the base of news content. To build an effective and useful fake news detection system, it is natural and necessary to explore auxiliary information from different perspectives.
The reasons for this change:
 a) it is often more timely and less expensive to consume news on social media compared with traditional news media, such as newspapers or television.
b) it is easier to further share, comment on it, and discuss the news with friends, relatives, or other readers on social media.
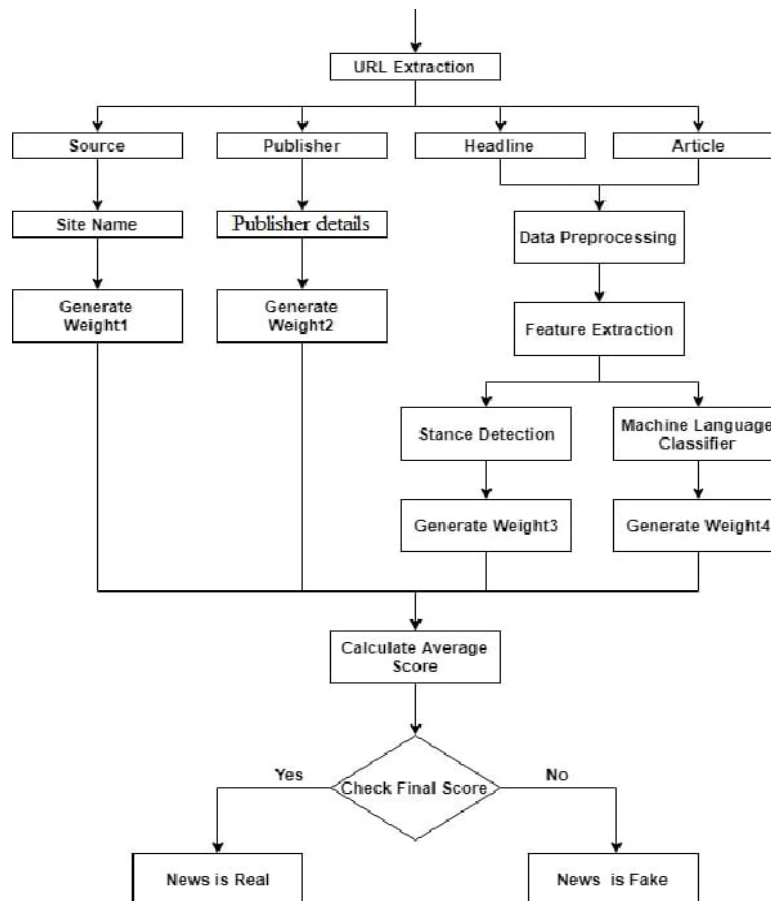
II.DEFINITION FOR FAKE NEWS

Fake news has existed for very long time nearly the same amount of time as news began to circulated after printing from the press. However, there was no exact definition for the word fake news. By that there was a lot of assumptions about this word literately. A narrow definition about this fake news was intentionally and verifiably false and mislead readers. While this news been published readers can't identify which news is real or false. By this misleading a lot of problems is been faced by the publisher, the main aim by this fake news was to mislead the readers and get money.

III.SIMPLE WORKING

Main working on this fake news detection is based on the URL extraction that means data's been extracted and analyzed to find the given news is real or fake.

As entering to the working section, it been mainly forced on the headline and article. On it first we are checking the condition of data preprocessing. That means we are removing the unwanted data links, dots, comma, pictures, space all will be removed. Then it been checking the condition of the feature extraction. That is, we are considering the words or headings that been counting to calculate how much times that repeated. Or while we are considering heading it is not necessary that the topic had been discussed. In this we are dividing the URL in to same parts so that its essay to calculate. This extraction is been divided in to stance detection and machine language classifier. That is how much time is it been repeated or sentence gap or frequency of that sentence. After all the weight been calculated.

The next working section is on publisher and source. In this section site name and publisher details is been discussed. And then we r generating the weight on it. After all we are calculated the average score on it. Check the final score, from that we can finalize the given URL is fake or real.

IV.EXISTING SYSTEM

◦   There are a lot of websites to detect or to find the given news is been real or fake.

◦   But I each website there will be only one or two type of implementation it been performed.

◦   In the existing system there are some limitations. That means sometimes it just calculate the ranking that been calculated by the system, it may be checking the protocol or matching of the contents.

V.PROPOSED SYSTEM

◦   Comparing to the existing system, the proposed system plays and important role on identify the aquatics result, that a news is been real or fake.

◦   In this proposed system we are implementing or comparing 5 major steps:

▪   Alexa ranking: It means of website popularity. By using this ranking system, we can rank millions of websites in order of its popularity. Alexa rank reveals how a website is doing relation to all other sites, which makes its great KPJ for benchmarking and competitive analysis.

▪   Google page ranking: It's also known as get google rank. It is an algorithm used to rank the web page that been googled in their search engine results. It is the way to measure the important of website page. According to google page

rank works by counting the numbers and quality of links to a page to determine a rough estimate of how important the website is.

▪ **Protocol:** It is the protocol that been established by the set of rules that determine how the data is been transmitted in between the devices of the same system. Essentially it allows connected devices to communicate with each other, regardless of a difference in their internal processes, structure or design. It's to identify the given URL is https or http.

▪ **Content matching:** Detecting the fake news in online media is been very important in our today's society as fresh news content is rapidly being produced false information to our society

▪ **Comparison:** As by using this comparison we can access the information result that the news been spreading though the social media. We can compare the content of the news that been published in another website. By that we can analyse the fact of news that been published in the other websites.

VI.WORKING OF FAKE NEWS

As we now all the working of this system is been on the URL. That is, we are checking out the news is been fake or real. By that URL is been received and parsed. So, while parsing this it been divided in to small pieces. After that we will be checking the protocol to make sure its https or http that means it's an extension of the hypertext transfer protocol. It is used for secure communication over a computer network and is widely used on the internet. After that we have to find the ranking of the given URL so to find it first, we have to determine the domain name. the function that been used is the getdomain_name. by this we are finding out the Alexa ranking and google page ranking. To find this both ranking we use the function Alexa rank and country ranging. In this we had been used natural language process that means we will be chosen some data to check it out whether if the same data been published in other websites. So, to check it out we use web scrolling. By this condition we will be checking of the content that been matched by the selected content and the tittle the bean similar to the tittle the been selected. Because in this fake news mainly we have we seen that content will not be matched the tittle that been displayed. By that itself the reader get wrong information about the news which be searched. This all conditions will be generating more aquarist news about been searched.

VII.ALGORITHMS THAT BEEN USED

TFIDF ALGORITHM
Machine learning algorithm cannot work raw text directly. Rather than text must be converted into vector of numbers. In this natural language a common technique used to extract the features from text is to place all of the words that occur in the text in a bucket. This approach is called a bag of words.
Term frequency: It is the number of times which the word been appears in a document that will be divided by the total number of words in the given document. Every document has its own term frequency.
Inverse data frequency: IDF the log of this number is to divided the document by the number of documents that contain the word w. it determines the weight of the rare words across all documents in the corpus.
ALEXA RANGING
Alexa ranking is a global ranking system that ranks millions of websites in order of popularity. It been calculated by looking at the estimated average daily unique visitors and number of page views for a given site over the past 3 months. Traffic and engagement are estimated from the browsing behavior of people in our global panel, which is sample of all internet users.
SIMILARITY SEARCH
It is a fundamental computational task that involves searching for similar items in a large collection of objects. It is the web crawling. It is also called spider or spiderwort it often used to check it out in other websites to find it out the publicity of the content or the aquarium of the content that been published under the title.

VIII.CONCLUSION

As we now day by day the popularity of social media in social media get increased and by that more and more people consume news from social media instead of traditional news media. So, by that the social media has been also used to spread fake news, which give a strong negative impact on the reader. By this we are losing the truth of the news that been published. We know that now a day's people trust in that they see first, it may be the title that been published. But the fact is that the tittle mayn't be the content tat been displayed. So, by this article we explain that while comparisons that been gone through 5 major steps will reduce the impact on this fake news. While going this we can analyze the fact in the news while been readied by the reader.

REFERENCE

https://www.diva-portal.org/smash/get/diva2:1341552/FULLTEXT01.pdf

https://digital.csic.es/handle/10261/192928

https://link.springer.com/chapter/10.1007/978-3-030-60975-7_17

https://paperswithcode.com/task/fake-news-detection

https://ieeexplore.ieee.org/document/8546944

https://www.hindawi.com/journals/complexity/2020/8885861/