International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Detecting Fake News

SUFIYA KAZI[1], GAYATRI BAJANTRI[2]

[1]Mtech (CNE), Assistant Prof(Mtech)
[2]SIET VIJAYPUR KARNATAKA(INDIA), SIET VIJAYPUR KARNATAKA(INDIA)

Abstract:- The proposed paper shows an approach with the aim of detecting fake statements made by public figures by use of artificial intelligence. Many different approaches were implemented as a software system and then tested against a data set of statements. Among all the results the best achieved result in binary classification problem (true or false statement) that is the correct or wrong is 86%. The results may be improved in several ways that are described in the article as well. Now days, the news information that we get it is not compulsion that the news is not always true. The Internet makes it harder to cross check the available information, because there are lot of sources that often even contradict and go against each other. All of this can be caused the emergence of fake news. The data from Mass media and social media have a great influence on us all the people. There are different sides that are interested in using this to achieve their political goals with the help of fake news. These methods provide false information the incorrect data in form of news to manipulate and fool people in many different ways.

Index terms: aim, approach, and fake, true, results, mass media.

1. INTRODUCTION

There is a saying that fake news problem may be solved automatically and easily, without any human interference in it, by use of artificial intelligence. This can be caused by the increase of deep learning technique and other artificial intelligence techniques showed us that they can be very effective and can be proved efficient in solving complex, difficult, lengthy and sometimes even the non-formal classification tasks.

Initially it was decided and planned to use only the statements collection of data themselves for the different classification purposes. This means that not even a single of the metadata available is used for classification. The classification algorithm can be further and can actually be improved in the upcoming future. The news channels provide false information in form of news and media to manipulate the people so that people belie in them easily and keep trusting them.

There are lots of websites and organizations with a single purpose and goal of spreading and giving out all only the false information. They publish fake news, false materials, hoaxes, conspiracy of the real news information. Examples of this false forwarding of information may be found in the countries such as Britain, Russia and Ukraine, United States of America, many other countries [1]. Thus, fake news is a global issue and problem and an important challenge to be tackled.

2. Proposed work

When the data and information is collected initially it is all divided further so that the work of detecting which data is incorrect becomes easy and successful. The classification algorithm that we are using for division of data is shown below step by step

The different steps that can be used for the pre-processing of collected information are the following:

- Dividing the statements into separate tokens (words).
- Removing all the numerical (numbers).
- Removing all grammatical punctuation marks.
- Delete all other non-alpha characters.
- the stemming procedure is to be applied to the rest of the remaining tokens
- Deleting all the stop words. Stop words are the words occur in basically all types of texts. These words are common and they do not really affect the meaning of the textual information.
- Substitution of words with code the code used here is with their tf-idf scores. In information retrieval for an example, tf–idf, which is a short for “term frequency–inverse document frequency. Everywhere when we want to use term frequency we can simply write tf automatically it will be substituted.

The figure 1 shows all the steps from login to log out to check our required data's accuracy.

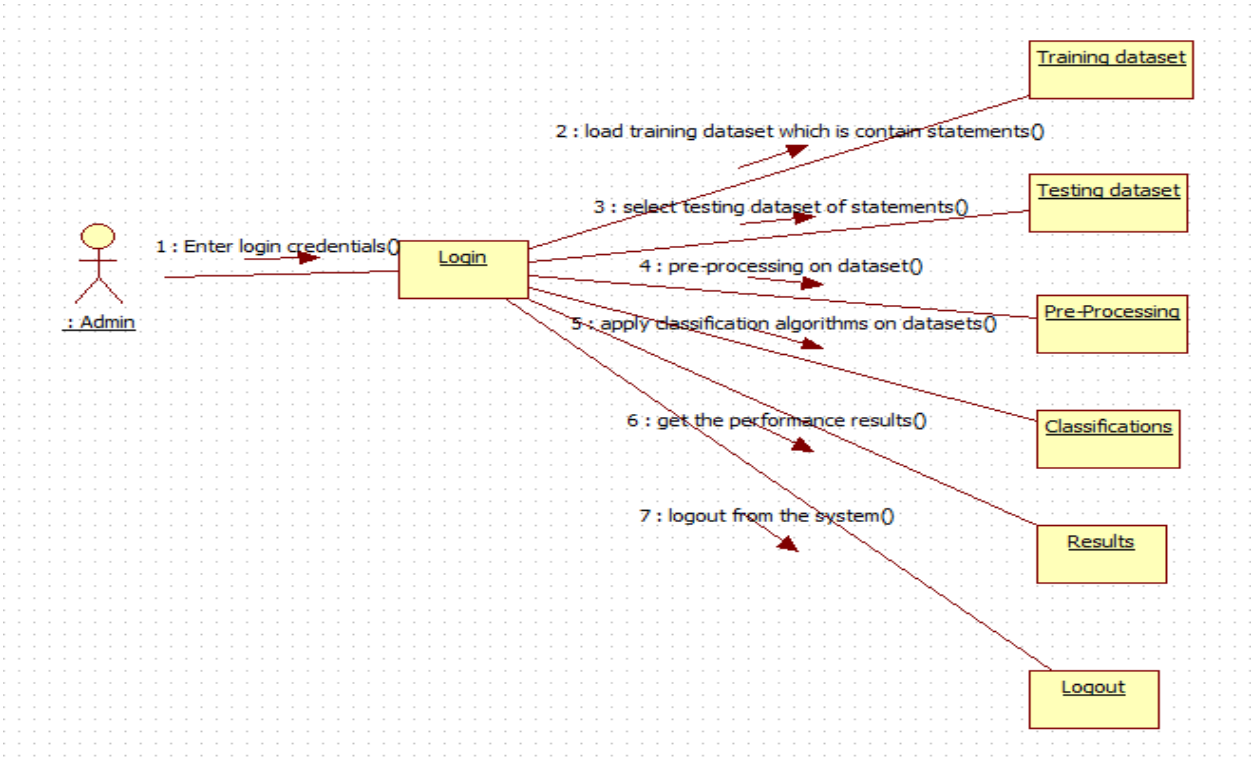

Figure 1: The working of project

There are three main types of algorithms used for this proposed work to check data and its accuracy that are given below, one by one each algorithm is taken and tested against set of data and every algorithm gives different output later which algorithm gives more effective result is compared and taken into considerations

Algorithms:

1. Naive Bayes
2. Support Vector Machines
3. Random Forrest Classifier

3. SYSTEM CONFIGURATION

Software requirements:

Operating System : Windows family (Like windows x,7,8,10)

Technology used : Python 3.6

IDE : PyCharm

Hardware requirements:

Processer : Any Update Processer

Ram : Min 4 GB and above it

Hard Disk of system : Min 100 GB

4. LITERATURE SURVEY

Deception detection for news: three types of fakes. In: reseedings of the 78th ASIS&T Annual Meeting: A fake news detection system aims to assist users in detecting and filtering out varieties of potentially deceptive news. The prediction of the chances that a particular news item is intentionally deceptive is based on the analysis of previously seen truthful and deceptive news. The analysis is only made when previous results are seen. The lack of deceptive news, available as corpora for predictive modeling, is a big block in this field of natural language processing (NLP) and deception detection. The proposed paper aims to discuss three different types of fake news, each in contrast and along with the to genuine serious reporting, and weighs their pros and cons their advantaged and disadvantages as a corpus for text analytics and predictive modeling works respectively. Filtering, vetting, and verifying online information continues to be important and useful in library and information science (LIS), as the gap between traditional news and online information are reducing day by day.

Further in the year of 2017. the work said as This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar and closet satire than real news". In: the 2nd International Works News and Public Opinion at ICWSM The problem of fake news has gained a lot of attention as it is claimed to have had a significant impact on 2016 US Presidential Elections. Fake news is not a new problem that has raised today and its spread in social networks is well-studied and well known by all of us. Often an underlying assumption and thinking is that in fake news discussion is that it is written to look like real news as if it is actual news that has come up, fooling the reader and people who believe them, who does not check for reliability and accuracy of the sources from where they get the information. Through a unique study of three different data sets and features that capture the style and the language of articles, we show that this assumption is not always true. Fake news in most of the cases is more similar to satire than to real news, leading us to conclude that persuasion in fake news is achieved through heuristics rather than the strength of arguments when studied. We show overall title structure and the use of proper nouns that can be utilized in titles that are very significant in differentiating fake from real the false from truth.

CONCLUSION

In the paper to be presented, several algorithms for classifying statements made by public figures were collected tested and implemented. Unsurprisingly and as everyone has knowledge, the best results and outputs were seen both in classification accuracy based, the classification accuracy which is further based on six categories and binary classification. This encourages and makes wider for future research with extensive usage of the deep neural networks. Achieved results might be significantly and gradually improved and better work can be seen further. It is possible to both improve the data the information which is used for training as well as the machine learning models themselves. And hence the data can be tested and real information and data will be reached to each of us.

References:

1. M. Granik, V. Mesyura, "Fake news detection using naive Bayesclassifier," 2017 IEEE First Ukraine Conference on Electrical andComputer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
2. Rajaraman, A.; Ullman, J. D. "Data Mining". (2011) Available:http://i.stanford.edu/~ullman/mmds/ch1.pdf. Accessed Mar. 24, 2018.
3. Stemming and lemmatization. (n.d.) Available:https://nlp.stanford.edu/IR-book/html/htmledition/stemming-andlemmatization-1.html. Accessed Mar. 24, 2018.
4. Sparck Jones, K. (1972). "A Statistical Interpretation of Term Specificityand Its Application in Retrieval".Journal of Documentation, vol 28, pp 11–21.
5. David A. Freedman (2009). “Statistical Models: Theory and Practice”. Cambridge University Press.p. 128.
6. Ho, Tin Kam (1995). “Random Decision Forests”. Proceedings of the3rd International Conference on Document Analysis and Recognition,Montreal, QC, 14–16 August 1995. pp. 278–282.
7. Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vectornetworks". Machine Learning, vol 20, pp 273–297.