International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Authentic News Summarization

Irfanali Shaikh[1], Nafis Borkar[2], DR. Shubhangi Vaikole[3]

[1,2]Student, Computer Department, Datta Meghe College of Engineering, Maharashtra, India
[3]Asso.Professor, Computer Department, Datta Meghe College of Engineering, Maharashtra, India

---

Abstract: *Implementation of Authentic News Summarization of news system has been discussed in this paper. The proposed system supports two major functionality namely* ***authentication*** *and summarization. The module of authenticity is given more priority over the summarization, therefore* ***authentication*** *of news leads to summarization of news such that if authenticity measure is greater than or equal to 55% only then the summarization module will execute or else system will return the authenticity parameters of news in terms of percentage. The problem of* ***fake news*** *must be addressed using* ***artificial intelligence*** *approaches and* ***machine learning*** *algorithm, Syntaxnet algorithm is used to extract phrases from user input* ***Statistical analysis*** *and keywords have been used in this system for verification of news and ensuring whether the given news article is a hoax news or not by comparing the contents from some authorized news website with the input news.*

Keywords — Authentication, Machine Learning, Artificial Intelligence, Fake news, Statistical Analysis.

1. INTRODUCTION

In this era of internet we all are aware of how unauthentic news has an impact on our daily life, when any news is encountered, we just forward it in exact same form without putting an additional effort of doing some research on it. Also few people having bad intention try to put more uncertainty in the news which may lead to intense hate toward some community or some organization such as political party to gain votes in an election. The proposed system can be used to reduce the potential threat of hoax news to a great extent, by using algorithms of artificial intelligence and machine learning. Today Mass media is easily accessible to every individual; therefore fake news reaches to a wide range of audience by spreading from one individual to another via any social media platform. Recent Google Trends analysis reveals that fake news came into picture in around the election time of US president in 2016, and it has remained popular since then. The fake news not only gains popularity by being attractive in nature but also it often leads the citizens being confused about their basic facts. A recent survey led by YouGov in 2017 addressed the audience with both hoax as well as the original news [1]. The individual able to differentiate between these news were very few, this highlights a potential threat to both the science and the society and it must be addressed by any artificial intelligence approach accordingly. The proposed system is not only capable of ensuring authenticity of the given news article but also it has a functionality of summarization, where the given news article is broken down into smaller and meaningful form, conserving time of the end users.

1.1 Present situation

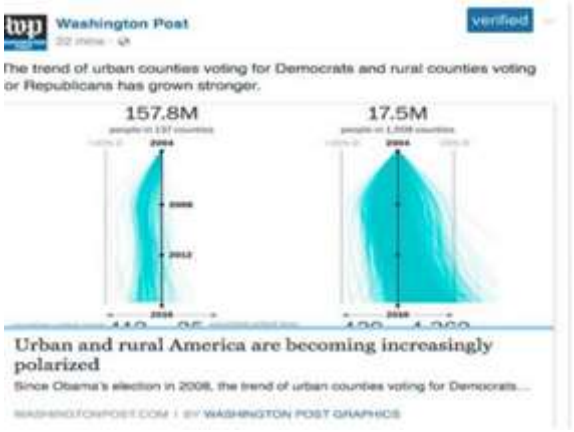

Fig 1(a): Original news

Fig 1(b): Hoax news

It seems clear that the decision of tracking the audience is given to the machine; even the commercial websites track the devices by means of cookies. Since the machinery is taking control over displaying news to the user, in many situation it becomes very difficult to identify some part of information is fact or not [2]. In figure 1 the top image is from verified source specifying votes in urban countries,

whereas there is hoax news of changes in climate due to increment of global warming, again this news was verified by the authorized people and reported as a fake.

2. LITERATURE REVIEW

In (Improving spam detection in Online Social Networks 2015), author proposed how the massive use online social media has increased nowadays, with great advancement comes great responsibility [3]. Similarly author classified users in two categories namely spammers and non-spammers the main objective of this system was to detect spammers in twitter Online social network. This is a very important aspect for our project! Techniques used to detect the spammer account present on a social media platform with accuracy 87.9% are:

Firstly the data is classified as spammer and non-spammer using learning algorithm such as.

I) Naive Bayes: given the attributes of user account the probability associated with the user of being a spammer was generated.

II) Clustering was unsupervised learning algorithm which can classify a given user into spammer/non-spammer.

III) Decision tree was also used for classification of user account in this method using the decision tree a decision was made at every level of the decision tree.

In (automatic text summarization: a reader oriented approach 1994), author proposed a new method for text summarization by specifying the informative and condensed nature of the given input which is a very important aspect for our system [4].

The method used in this system is called ROSE- Reader Oriented Summary Engine; the primary goal of this engine was to generate the summary from the perspective of user rather than generating summary from perspective of text.

In (Spotting Fake News: A Social Argumentation Framework for Scrutinizing Alternative Facts) author proposes a prototype to verify the proposed alternative fact to reduce the proliferation of fake news by describes the impact of fake news over the internet [5].

Measures were taken in order to help users to verify the alternative facts, most of the time fake news emerges through improper knowledge of subject base. Every other system verifies the news with some ground knowledge of truth whereas this system was designed to match the most relevant news with the prescribed news detail. Approaches such as analyzing the conclusion, the quality of arguments are judged and deciding portions of issue to be analyzed. This system not only had the ability to detect the fake news also provides the statistical analysis to user to analyze future fake news article.

In (Challenges in automatic summarization) author provides an insight of all the problems involved in designing an effective text summarization technique. With development alongside the researchers are investigating and implementing summarization tools to extract meaningful information from the natural language input to a machine. The nature of the summary can differ from one algorithm to another. The reason for this difference is associated in the ground level, whether they are extractive or abstractive in nature.

3. EASE OF USE

Fig2(a): Web page (user interface of the system)

In order to remove the complexity of the overall execution we designed a website using HTML, CSS and Bootstrap. Again to increase the overall functionality Django provided hassle free environment and clean pragmatic design of web frame work. The website is very simple in nature, taking the raw news data (URL or any news headline) and generating the results ensuring whether the given news was true or false associating the probability of truth.

Fig2(b): Web page (user interface of the system)

3.1. USER INTERFACE

The web page is separated by divisions each division has given a unique id and the entire functionality is built on a single web page to minimize the conflict while using the web page, second division is where the main functionality is implemented allowing user to enter the news for further verification the text limit of the system is 200 words for once. Therefore the actual complexity of executing high end machine leaning algorithm is abstracted from the user, Providing a simple and user free environment to the user.

4. EXISTING SYSTEM

There are many authenticating algorithm developed to address the problem of fake news, Whereas the algorithm can be applied on back end of the system, therefore to use the algorithm one must be technically strong to use concepts such as python, anaconda. 'Identifying tweets with fake news' was designed by Saranya Krishnan, Min Chen. The system was based on techniques such as statistical analysis of user's account, reverse image searching and cross verification [6]. Also a 'Word Sequence Models for Single Text Summarization' was implemented by Yulia Ledeneva and Rene Amulfo Garcia Hemandez, techniques such as n-graphs and maximal frequent word were used in order to extract text summarization. Our system contains of two modules authentication and summarization.

5. PROPOSED SYSTEM

The main objective of this system is to verify the given news using mechanism such as cross-verification of news from various sources.

The system is being implemented in python 3.6 version or higher for more convenience anaconda was installed as it provides an easy environment for python execution [7].

Packages such as Sklearn, Numpy and scipy were installed. Sklearn provides basic mechanism for execution of machine learning algorithm such as classification, regression and clustering for implementation of news authentication system [8]. Numpy provides a set of tools associated with a multidimensional array used for scientific computing in python [9]. Scipy was installed for enabling mechanism such as data processing [10]. It provides environment such as MATLAB, Octave and Scilab[11].

The system takes news data in form of text as an input and selects keywords based upon the confidence score associated with it [12]. Algorithm such as Syntaxnet is used, this algorithm is famous and designed and is open source neural network framework. Syntaxnet algorithm provides mechanism to extract meaningful data from natural language, for processing human language in order to perform intelligent operation [13]. Syntaxnet is mainly used in areas such as feature extraction, representing annotated data, and evaluation.

At initial the system request data in text format and provide acknowledgement of the input along with returning the nature of the news such as true or false. The system was also able to return the confidence score and truth probability associated with the input news.

Fig 3(a): Initial Execution

The above figure describes actual initial phase execution of the system. At initial we kept our focus on Authentication module .The complexity of program remains hidden from the user as the program execution is at the back end, the user will interact with the system through a website discussed in ease of use section providing convenient accessibility to the end user.

Fig3 (b): Progressive Execution 1

Fig3(c): Progressive Execution 2

Necessary changes were made to optimize the execution of the program such as updating in the extracting method to limit the word count being compared from external sources to authenticate the given news. The Every keyword was processed to compute its relevance score, minimum support count was set to 3 and Keywords present in the

input were extracted using phase extractor along with syntaxnet algorithm, the output consists of highlighted keywords along with their confidence score and true or false nature of news in terms of percentage.

6. SYSTEM DESIGN AND ARCHITECTURE

Detailed Architecture of proposed system is given below. The system can be classified in two major categories authentication and summarization. The initial step in the system is to enter the news intended by the user to verify. This input is then processed to achieve desire result (Nature of the news). Accordingly summarization will take place Depending on authenticity of news.

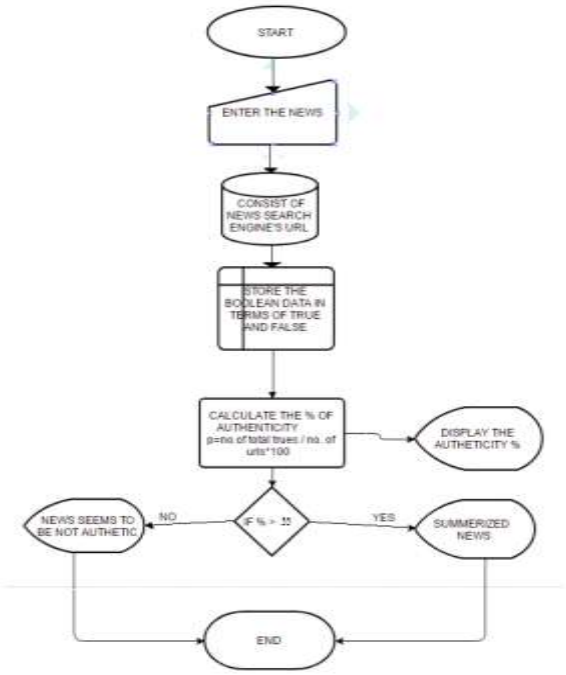

Fig 4: System architecture

The actual data flow begins with the user's input to the system. This input is fed to the proposed system to return authenticity parameters of news in terms of percentage. For summarization of the input news machine learning algorithms such as syntax Syntaxnet. Syntaxnet algorithm provides mechanism to extract meaningful data from natural language, for processing human language in order to perform intelligent operation [13]. The system has a threshold value if the input passes the threshold then only the given news will be returned in summarized manner or else the given news will be classified as unauthentic and truth probability associated with the news will be reflected to user.

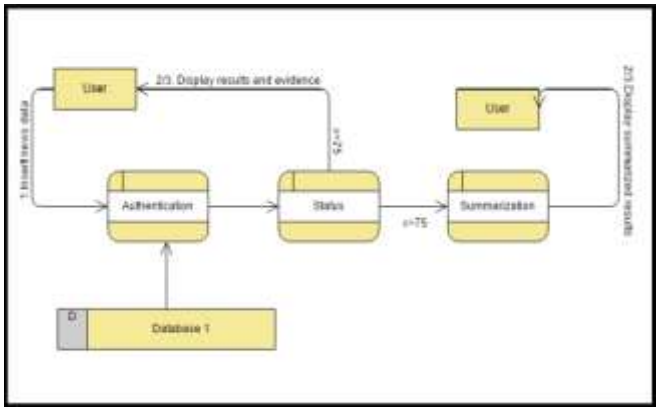

Fig 5: Data Flow diagram

7. CONCLUSIONS

Proliferation of online social media has resulted in widespread of hoax news as it is easily transferred from one user to another by any social media platform. By studying this system it can be concluded that.

- Awareness is very necessary as a responsible citizen in order to maintain integrity and conflict free environment related to daily life.
- The system was able to achieve 65% of accuracy in verifying news also it provided with statistical data which will help user to identify future fake news.
- Syntaxnet algorithm provided a great mechanism to extract meaningful phrases from natural language allowing the summarization of news if authenticity parameters are more than 55%.

8. ACKNOWLEDGEMENT

This research was supported by Datta Meghe College of Engineering. We thanks our project guide Dr.Shubhangi Vaikole who provided advisory insight and expertise along with great support and resources in our research

9. REFERENCES

[1]Yougov survey conducted on 2017: https: //yougov.co.uk/topics/politics/articles-reports/2017/03/23/what-counts-fake-news

[2]Information of present situation related to fake news situation :https://www.researchgate.net/publication/321820496_The_current_state_of_fake_news_challenges_and_opportunities

[3] Improving spam detection in Online Social Networks 2015:https://ieeexplore.ieee.org/document/7100738

[4] automatic text summarization: a reader oriented approach
1994:https://ieeexplore.ieee.org/document/397011

[5]Spotting Fake News: A Social Argumentation Framework
Facts:https://ieeexplore.ieee.org/document/8029851

[6]Identifying and analyzing the tweets with fake news: https://ieeexplore.ieee.org/abstract/document/8424744

[7] Minimum requirements for implementing proposed system:https://repo.continuum.io/pkgs/

[8] Details about scikit package and mechanism provided by scikit:https://scikit-learn.org/stable/

[9]Details of Numpy providing various set of tools: https://www.geeksforgeeks.org/numpy-in-python-set-1-introduction/

[10] Determining andDifferentiating between Numpy and Scipy:https://www.quora.com/What-is-the-difference-between-NumPy-and-SciPy

[11] Information of mechanism used in detail regarding Scipy: https://www.journaldev.com/18106/python-scipy-tutorial

[12]Summarization module:
https://www.paralleldots.com/keyword-extractor

[13]Information about syntaxnet package:https://github.com/tensorflow/models/tree/master/research/syntaxnet