International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Fake News Detection

Aman Gautam[1], Sachin Yadav[2], Rishi Kataria[3], Prof. Mrunali Desai[4]

[1,2,3]Student, Dept. of Computer Engineering, K.J. Somaiya Institute of Engineering & I.T, Sion, Maharashtra, India
[4]Professor, Dept. of Computer Engineering, K.J. Somaiya Institute of Engineering & I.T, Sion, Maharashtra, India

---

Abstract - *Social media acts as a double-edged sword ie. Its low cost as well as spread over the globe for consumption of news through social media. The rapid spread of misleading information in media outlets such as social media feeds, news blogs, and online newspapers have made it challenging to identify trustworthy news sources, thus increasing the need for tools able to provide insights into the authenticity of online content. This paper aims to present an insight into the characterization of the news story in the modern diaspora. Subsequently, we dive into fake news detection approaches that are based on the text-based analysis. The method for all kind of Fake news which is circulated over the Internet and obtaining the result in form of percentage accuracy is implemented.*

Key Words: Biometric, Facial Recognition, Fingerprint Module, Raspberry Pi, LBPH Face Recognizer, Haar Cascade, Firebase, Customized report

1. INTRODUCTION

An increasing amount of time from our lives is spent on social media, interacting online through social media platform which further increases the consumption of news from social media rather than the conventional news sources such as News Papers. The reasons for this change in news consumption behaviour are

(i) it is more timely and less expensive to consume news on social media compared with traditional news media, such as newspapers and television.

(ii) It is easier to share, post, comment and discuss with friends over social media.

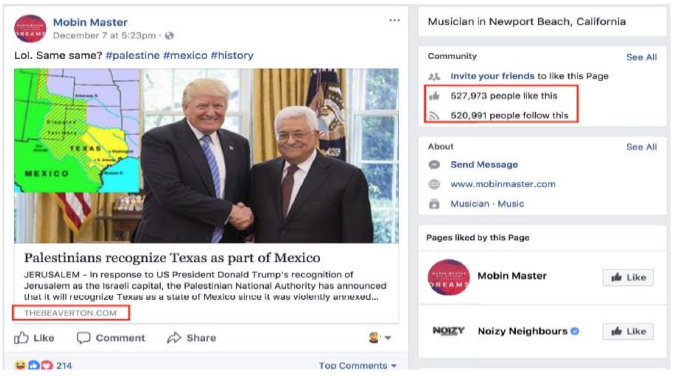

Figure 1: An interpretation of how the story titled "Palestinians recognizes Texas as part of Mexico" appears on Facebook

[Source: http://www.facebook.com/]

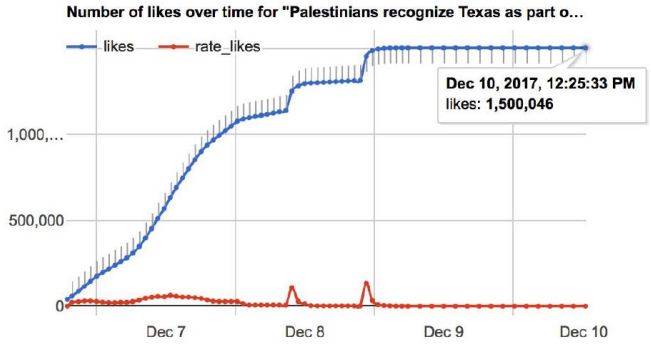

Figure 2: Facebook Trend Report on "Palestinians recognizes Texas as part of Mexico"
[Source:http://www.trendolizer.com/]

There are various motivations behind spreading Fake news are making political gains, harming the reputation of businesses, as click bait for increasing advertising revenue, and for seeking attention, spreading unrest within a community.

Fake news detection focuses on identifying the percentage of unreliable data within the news. Many techniques have been applied to fake news detection such as Latent Semantic analysis, Singular value decomposition, recurrent neural network, Deep Learning.

II. LITERATURE SURVEY

TYPE OF DATA IN NEWS

In this section, we have a tendency to discuss the categories of knowledge that the news stories are fabricated from; there are four major formats during which users consume their news. Some can be additional common than alternative types, however they're all major varieties.

1) *Text:* Text/string content is usually analyzed by text-linguistic and it's a branch of linguistics, that in the main focuses on the text as a communication system. It is much over simply sentence and words, it has characteristics like tone, grammar, and has linguistics that allows discourse analysis.

2) *Multimedia*: rather like name defines, it's associate degree integration of multiple types of media. This includes pictures, video, audio, and graphics. this is often terribly visual and catches viewers attention at terribly initial.

3) *Hyperlinks or Embedded Content*: Hyperlinks modify writers to link off to completely different sources and gains readers trust by proving the hypothesis of the news article. With advent of social media, writers tend to insert a snapshot of relevant social media post (e.g., Facebook post, tweet, YouTube video, soundcloud clip, Ingram post, etc.)

4) *Audio*: Audio may be a a part of transmission class, but it has standalone medium to be a news supply. This category includes podcast, broadcast network, radio service and this medium reaches bent on the larger audience to deliver the news

TYPES OF FAKE NEWS

Social science researchers have studied fake news from different perspectives and provided a general categorization of different types of fake news.

(i) *Visual-Based :* This consists of misleading graphical representations such as photoshopped images ,videos etc.

(ii) *User-Based :* Such types of news are for a particular group of people which can be classified by age , gender , culture.

One such example is of Blue Whale game which targeted only kids.

(iii) *Post-based :* These fake news appear on social media platforms which basically contains images or videos with caption , tweet and memes attached to them.

(iv) *Knowledge-based :* these are the news which provide scientific and reasonable explanation for unresolved issues which mislead the audience , e.g. false article to cure cancer.

(v) *Style-based* : Style-based focuses on the way of presenting to its readers, fake news are written by majority of people who are not journalists - that being said the style of writing can be different.

The above five are generalistic classification of Fake news around the globe.

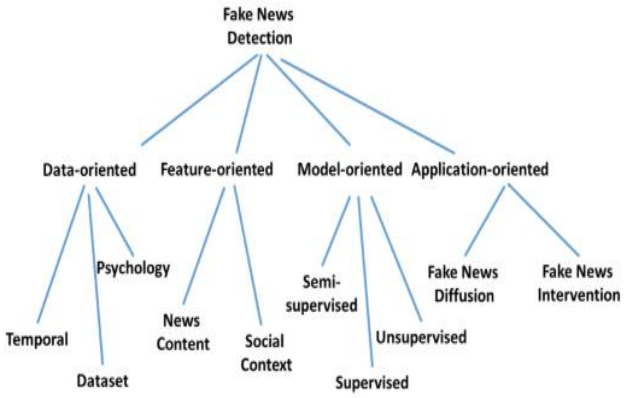

Figure 3: Various methods to identify fake news[Source:https://www.kdnuggets]

The above figure shows some of the the methods which could be used to identify a news as fake news and the below are the detailed versions of the above methods.

(A) Clustering based Methods

Clustering is a known method to compare and contrast a large amount of data, in, authors have used Graphical CLUstering TOolkit clustering package to help differentiate similarity between reports based on clustering algorithm[4]. This method involves running a large number of data set and forming/sorting a small number of clusters using Hierarchical clustering with the k-nearest neighbor method, clustering similar news reports based on the normalized frequency of relationships.

The ability to detect the misleading value of a new story is measured based on the principle of coordinate distances calculated. According to what the author claims based on achieving 63% of success using this method seems to be very useful on large datasets. One challenge could be that this approach might not be able to provide the accurate result if it is applied to a very recent fake news because the similar news story sets might not be available ie. lack in availability of similar Data.

(B)Predictive Modelling based Methods

The authors proposed a logistic regression model based on a training data set of 100 out of 132 news reports. According to this approach, positive coefficients increase the probability of truth and negative one increase the probability of deception. 70% of accuracy on training data-set and 56% of accuracy on test data-set can be achieved by this method[6]. Authors claimed that regression indicators like Disjunction, Purpose, Restatement, and Solution hood points to truth, and Condition regression indicator pointed to deception [6]. It is very important to note that both Clustering and Predictive Modeling has a success rate of 63% and 70%

respectively. However, Predictive Modeling approach shows real promise to perform instant fake detection, machine learning techniques can be used to improve the coefficients in an ongoing way.

(C) Content Cues based Methods

This method is based on what journalists like to write for users and what type of News user like to read (Choice Gap). These are contaminated News Stories which attracts the users. Such type of News can be produced by different sources resulting in multiple versions of the same News.

(i) Lexical and Semantic Levels of Analysis

Choice of vocabulary(synonyms and antonyms) play important role in convincing the user about truthness of the News Story.

(ii)Syntactic and Pragmatic Levels of Analysis

This is done by making reference to forthcoming parts in the news story.

Headlines are written to fill empty thoughts with leveraging ensuing text[5].This analysis also covers measuring news sites which have more share activity compared to sites that substantially produces more news content.

(D) Linguistic Features based Methods

- Ngrams: Term Frequency inverse document frequency (TFIDF) table stores the data to reflect how important a word is to the document that it is used in.
- Punctuation: Use of punctuation can help the fake news detection algorithm to differentiate between deceptive and truthful texts. Punctuation feature collects eleven forms of punctuation, that is enforced through this detection.
- Psycho-linguistic features: It gives the relation between linguistic factors and psychological aspects. LIWC(Linguistic enquiry and word count) picks out appropriate proportions of words[2]. This allows the system to determine the tone of the language (e.g., positive emotions, perceptual processes), statistics of the text (e.g., word counts), part-of-speech category (e.g., articles, verbs).
- Readability: This includes extraction of content features such as the number of characters, complex words, long words, number of syllables, word types, and number of paragraphs[2]. Having these content features allow us to perform readability metrics, such as Flesch-Kincaid, Flesch Reading Ease, Gunning Fog, and the Automatic Readability Index (ARI).

IV. Implementation

To develop an hybrid application which will take care of the fake news problem by giving the probability of the news being authentic. A system which will help users identify whether the news being circulated on social media such as facebook, whatsapp is fake or not and helping the users by providing them the authenticity of news.

Our approach first summarize the fake news using text Rank.

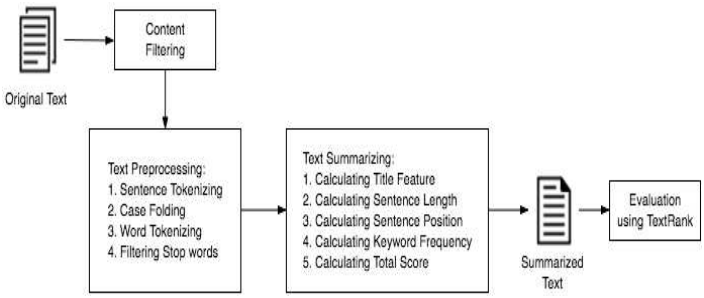

Figure 4: Summarization of text[Source:https://picswe.com]

At next step Web Crawling will be performed through the internet to find related article to the target text.

Web crawling will identify the similar articles to the target news and will extract all the articles. Now all the articles are compared to each other using LSA.

V. LATENT SEMANTIC ANALYSIS (LSA)

Now all the articles stored in database after web crawling and the target News will get compared with the help of Latent Semantic analysis and output would be a Cosine Distance Matrix which would give the similarity between the articles and target news.

In case Data is not similar still there is a scope to find connection between news then WMD (Word Movers Distance) will find the similarity in case data is different.

Now Fuzzy logic will be used to map the score of of LSA and WMD to single score . This score would be the final score to identify the authenticity of the target news.

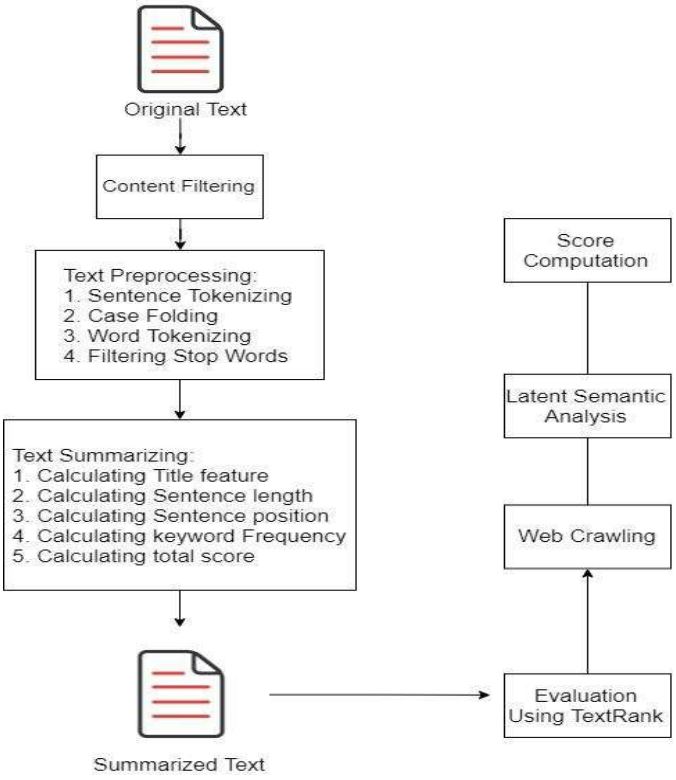

Figure 4: Flowchart

The above figure summarizes the implementation of Fake news Detection.

VI. RESULTS

Figure 5: Output Screen

VII. CONCLUSION

In this review paper, we addressed the task of automatic identification of fake news. We introduced new fake news dataset , obtained by covering six news domains and developed classification models that rely on a combination of lexical, syntactic, and semantic information, as well features representing text readability properties. Our best performing models achieved accuracies that are comparable to human ability to spot fake content.

REFERENCES

[1] Eugenio Tacchini, Gabriele Ballarin, Marco L. Della Vedova, Stefano Moret, and Luca de Alfaro, Some Like it Hoax-Automated Fake News Detection in Social Networks,Technical Report UCSC-SOE-17-05 School of Engineering, University of California, Santa Cruz

[2]Veronica Perez-Rosas , Bennett Kleinberg , Alexandra Lefevre, Rada Mihalcea, Automatic Detection of Fake News Computer Science and Engineering, University of Michigan Department of Psychology, University of Amsterdam

[3] Zhixuan Zhou, Huan Kang Guan, Meghana Moorthy Bhat and Justin Hsu Fake News Detection via NLP is Vulnerable to Adversarial Attacks, Hongyi Honor College, Wuhan University, Wuhan, China Department of Computer Science, University of Wisconsin-Madison, Madison, USA

[4] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang , and Huan Liu Fake News Detection on Social Media: A Data Mining Perspective, Computer Science & Engineering, Arizona State University, Tempe, AZ, USA Charles River Analytics, Cambridge, MA, USA Computer Science & Engineering, Michigan State University, East Lansing, MI, USA

[5]Shivam B. Parikh and Pradeep K. Atrey, Media-Rich Fake News Detection: A Survey, Albany Lab for Privacy and Security, College of Engineering and Applied Sciences, University at Albany, State University of New York, Albany, NY, USA

[6] Automatic Online Fake News Detection Combining Content and Social Signals

THE 22ND CONFERENCE OF FRUCT ASSOCIATION