

FAKE NEWS DETECTION USING MACHINE LEARNING

¹ Ms.CH.UMA DEVI, ² R.PRIYANKA, ³ P.S.SURENDRA, ⁴ B.S.PRIYANKA ⁵ CH.N.D.L.NIKHILA,
¹ Assistant Professor^{2,3,4,5}UG Student, , Department of Computer Science and Engineering,
Godavari Institute of Engineering & Technology, Rajahmundry, AP

Abstract: A lot of things read online, especially in the social media feeds, may appear to be true, often is not. Fake news is news, stories or hoaxes created to deliberately mislead or deceive readers. Usually, these stories are created to either influence people's views, push a political agenda or cause confusion and can often be a profitable business for online publishers. The dissemination of fake news in today's digital world has effected beyond a specific group. Mixing both believable and unbelievable information on social media has made the confusion of truth. That is, the truth will be hardly classified. This paper comes up with the applications of NLP (Natural Language Processing) techniques for discerning the 'fake news', which is the misguiding news that is being published through unknown sources.

Keywords: NLP, logistic regression, naïve bayes classifier, SVM

1. INTRODUCTION

These days, fake news is creating different issues from sarcastic articles to a fabricated news and plan government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is "fake news" but lately blathering social media's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints.

The importance of disinformation within American political discourse was the subject of weighty attention, particularly following the American president election. The term 'fake news' became common parlance for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page views. In this paper, it is seeked to produce a model that can accurately predict the likelihood that a given article is fake news. Facebook has been at the epicenter of much critique following media attention. They have already implemented a feature to flag fake news on the site when a user sees it; they have also said publicly they are working on to distinguish these articles in an automated way. Certainly, it is not an easy task. A given algorithm must be politically unbiased – since fake news exists on both ends of the spectrum – and also give equal balance to legitimate news sources on either end of the spectrum. In addition, the question of legitimacy is a difficult one. However, in order to solve this problem, it is necessary to have an understanding on what Fake News is. Later, it is needed to look into how the techniques in the fields of machine learning, natural language processing help us to detect fake news.

The main purpose of this system is to detect the fake news, which is a classic text classification problem with a straight forward proposition. It is needed to build a model that can differentiate between "Real" news and "Fake" news.

2. EARLY HISTORY

The credibility of information was defined by many words such as trustworthiness, believability, reliability, accuracy, fairness, objectivity, and other with the same concepts and definitions [1]. There are several researches that use the machine learning approach to calculate the credibility of message [2][3][4]. Fake news is the contents that claim people to believe with the falsification, sometimes it is the sensitive messages. When the messages were received, they will rapidly disperse it to other. The dissemination of fake news in today's digital world has affected beyond a specific group. Mixing both believable and unbelievable information on social media has made the confusion of truth. That is the truth will be hardly classified. However, the appearance of fake news causes great threat on the safety of people's lives and property. There is misinformation (the distributor believes there are true) or disinformation (the distributor knows it is not fact but he intentionally hoax) in fake news proliferation. In this paper, we develop computational resources and models for the task of fake news detection. We present the construction of a novel dataset covering two different domains. The dataset is collected using a combination of manual and crowdsourced annotation efforts. Using this dataset, we conduct several exploratory analyses to identify linguistic properties that are predominantly present in fake content, and we build fake news detectors relying on linguistic features that achieve accuracies of up to 78%. To place our results in perspective, we also compare the accuracy of our fake news detection models with an empirical human baseline accuracy.

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts. Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature.

Conroy, Rubin, and Chen outline several approaches that seem promising towards the aim of perfectly classify the misleading articles [5]. They note that simple content-related n-grams and shallow parts-of-speech (POS) tagging have proven insufficient for the classification task, often failing to account for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars (PCFG) have been shown to be particularly valuable in combination with n-gram methods. Feng, Banerjee, and Choi are able to achieve 85%-91% accuracy in deception related classification tasks using online review corpora [6].

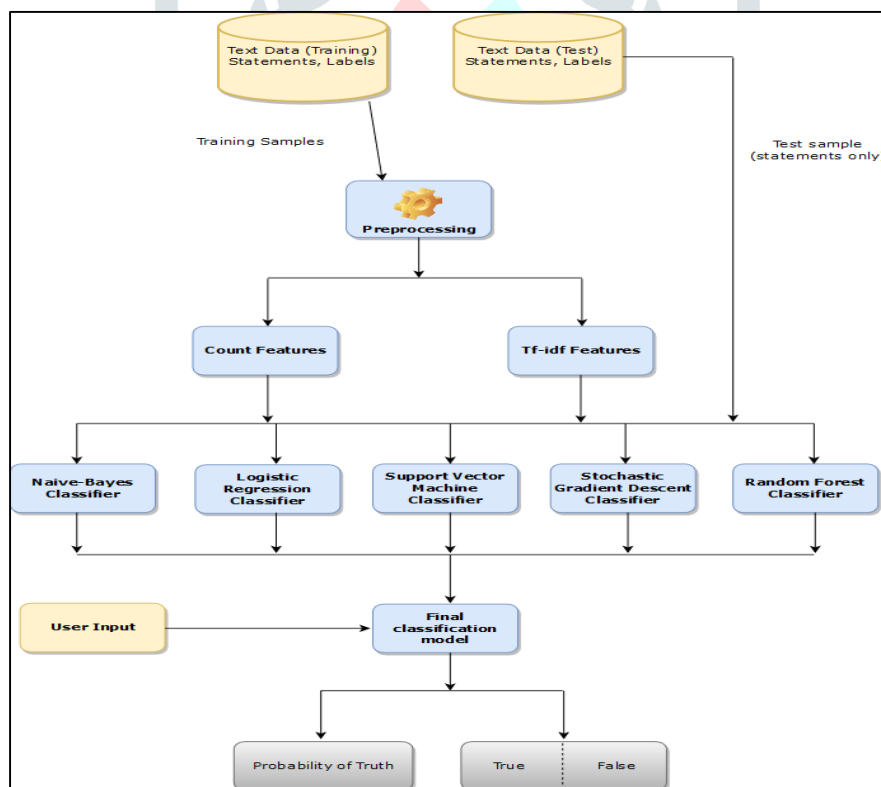
These models do not consider the important qualities like word ordering and context. It is very possible that two articles that are similar in their word count will be completely different in their meaning.

3. METHODOLOGY

This paper comes up with the applications of NLP (Natural Language Processing) techniques for detecting the 'fake news', that is misleading news stories that comes from the non-reputable sources.

In this paper a model is built based on the count vectorizer or a tfidf matrix (i.e. word tallies relatives to how often they are used in other articles in your dataset) can help. Since this problem is a kind of text classification, implementing a Naive Bayes classifier and Logistic Regression will be better as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf vectorizer) and choosing which type of text to use (headlines vs full text). Now the next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as "the", "when", and "there" and only using those words that appear at least a given number of times in a given text dataset.

4. SYSTEM ARCHITECTURE



Initially, the training data is extracted and the TFIDF matrix & count vectorizer are generated after preprocessing. This preprocessed data and the test data are sent to various classifier methods. A final classification model is selected. In this paper, logistic regression classification model is used. The input data is sent to the final classification model which gives result of the test data i.e., "True" or "False" and also gives the probability of truth as output. This is clearly explained by the above figure.

5. DATASET DESIGN

We started by collecting a dataset of legitimate news belonging to six different domains (sports, business, entertainment, politics, technology, and education). The news were obtained from a variety of mainstream news websites (predominantly in the US) such as the ABC News, CNN, USA Today, NewYorkTimes, Fox News, Bloomberg, and CNET among others.

To ensure the veracity of the news, we conducted manual fact-checking on the news content, which included verifying the news source and cross-referencing information among several sources. Using this approach, we collected 40 news in each of the six domains, for a total of 240 legitimate news.

BUILDING CROWDSOURCED DATASET

Collecting Fake News using Crowdsourcing - To generate fake versions of the news in the legitimate news dataset, we make use of crowd-sourcing via Amazon Mechanical Turk, which has been successfully used in the past for collecting deception data on several domains, including opinion reviews (Ott et al., 2011b), and controversial topics such as abortion and death penalty (Pérez-Rosas and Mihalcea, 2015).

However, collecting deceptive data via AMT poses additional challenges on the news domain. First, the reporting language used by journalists might differ from AMT workers language (e.g., journalistic vs. informal style). Second, journalistic articles are usually lengthier than consumer reviews and opinions, thus increasing the difficulty of the task for AMT workers as they would be required to read a full news article and create a fake version from it.

To address the former, we asked the workers to the extent possible to emulate a journalistic style in their writing. This decision was motivated by the 5th point of the fake news corpus guidelines described in section 3, which suggests to obtain news with homogeneous writing style. To address the latter, we opted to working with smaller information units. Our approach consists of manually selecting a news excerpt that briefly describes the news article.⁴ Thus, from the legitimate news dataset collected earlier, we manually extracted 240 news excerpts. The final dataset consists of 33,378 words. Each news excerpt has on average 139 words and approximately 5 sentences.

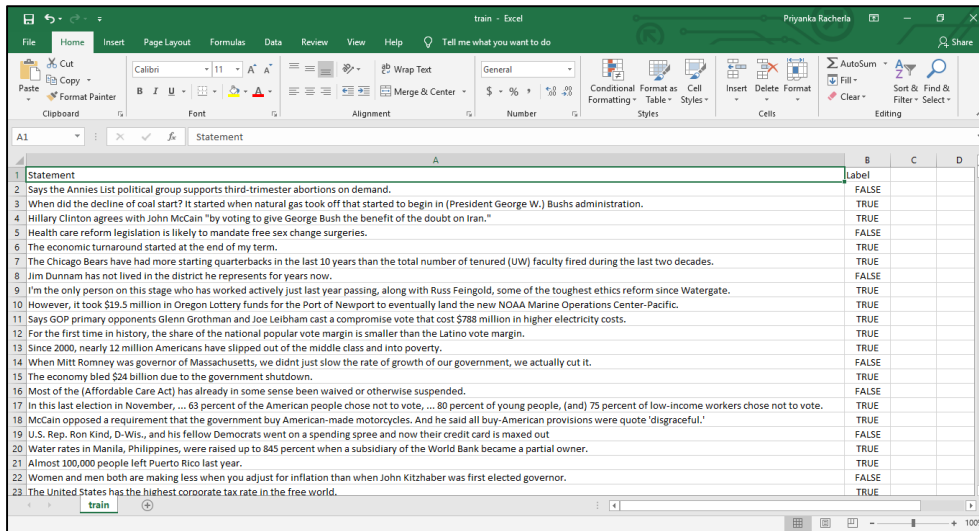
BUILDING A WEB DATASET:

A second dataset of fake news is collected from web sources following similar guidelines as in the previous dataset. However, this time, we aimed to identify fake content that naturally occurs on the web. We opted for collecting news from public figures as they are frequently targeted by rumors, hoaxes, and fake reports. We focused mainly on celebrities (actors, singers, socialites, and politicians) and our sources include online magazines such as Entertainment Weekly, People Magazine, Radar Online, among other tabloid and entertainment-oriented publications. The data were collected in pairs, with one article being legitimate and the other fake. In order to determine if a given celebrity news was legitimate or not, the claims made in the article were evaluated using gossip-checking sites such as “GossipCop.com”, and were cross-referenced with information from other sources.

During the initial stages of the data collection, we noticed that celebrity news tends to center on sensational topics that sources believe readers want to read about, such as divorces, pregnancies, and fights. Consequently, celebrity news tends to follow certain celebrities more than others further leading to an inherent lack in topic diversity in celebrity news. To address this issue, we evaluated several sources to make sure we obtain a diversified pool of celebrities and topics. Upon beginning the data collection procedure using these guidelines, another characteristic surfaced: several pairs contained nearly the same information with similar lexicon and reporting style, with differences being as simple as just negating the false news. For example, the following headlines correspond to a news pair where the legitimate version only negates the fake version: “Aniston gets into fight with husband” (fake) and “Aniston did NOT get into fight with husband” (legitimate). To address this issue, we sought to identify related news that still followed the fake-legitimate pair property while being sufficiently diverse in lexicon and tone. In the former example, the fake news was paired with an article titled “Aniston and Husband enjoy dinner” that was published on the date of the alleged fight.

SAMPLE DATASET:

The required dataset is represented in the excel sheet with two columns named statement and label. The news headline is mentioned in the statement column and the result stating true or false is placed in the label column. The sample excel sheet is mentioned below in the fig below.



6. SYSTEM IMPLEMENTATION

This system uses three methods to classify the believable and unbelievable message. They are Naïve Bayes, Logistic regression, and Support Vector Machine (SVM).

NAIVE BAYES: Naïve Bayes is the well-known classification method. We define the collected tweet data T and class of data (Cx) which x are believable and unbelievable. The probability of tweet data T in the class Cx can calculate as follow [7]:

$$P(Cx|T) = P(T|Cx) \times P(Cx) / P(T) \quad (1)$$

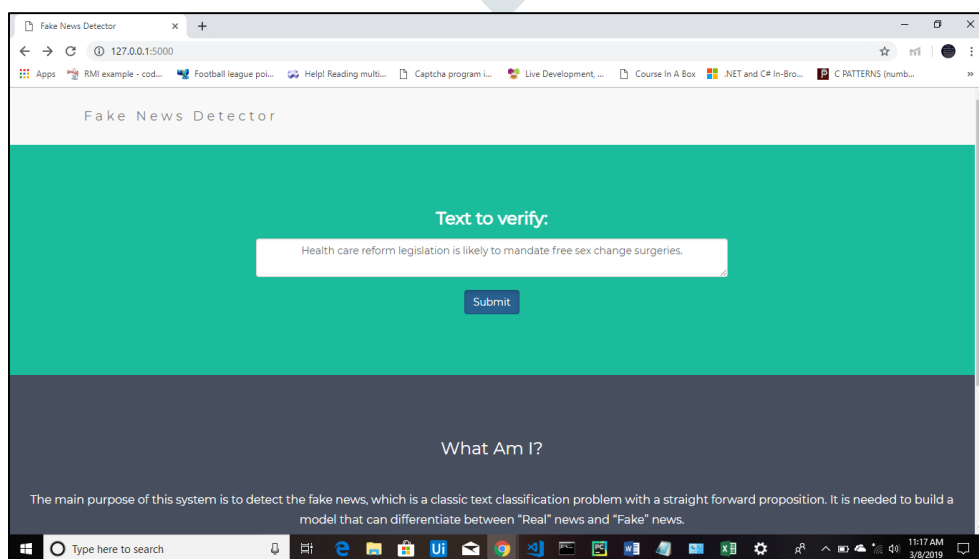
LOGISTIC REGRESSION: Logistic regression is another technique borrowed by machine learning from the field of statistics. It is the go-to method for binary classification problems (problems with two class values)

SUPPORT VECTOR MACHINE: Support Vector Machine (SVM) is the classification method of supervised learning. There uses the hyperplane to splits two data class's point with the maximum margin.

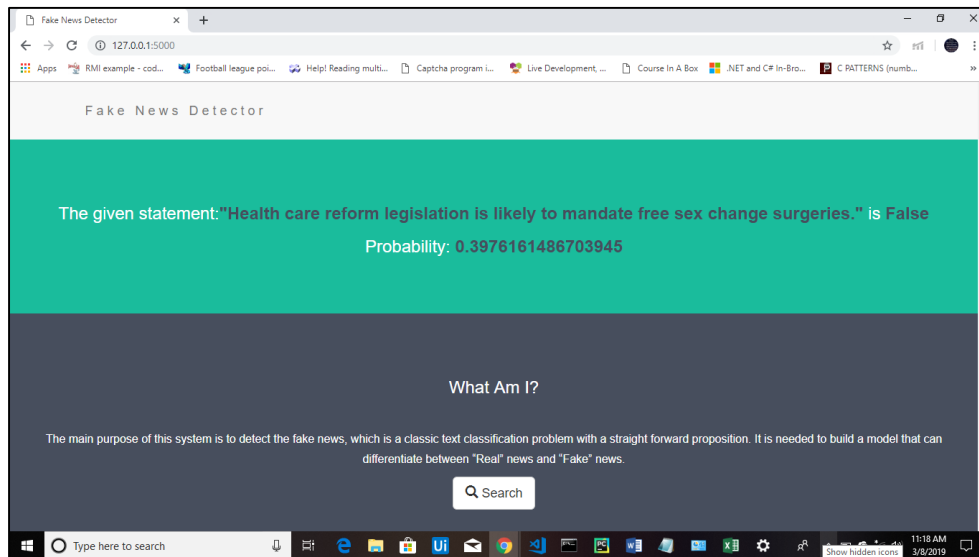
There are four evaluation results. They are precision, recall, Fmeasure and accuracy which are computed from True Positive, True Negative, False Positive, False Negative.

7. RESULTS

The news statement which is to be tested is entered in the textbox that is present under the caption "Text to verify:" as shown below:



The input data which is entered is processed through the proposed methodology and displays the output results as follows:



8. CONCLUSION AND FUTURE SCOPE

In this paper, we addressed the task of automatic identification of fake news. We introduced two new fake news datasets, one obtained through crowdsourcing and another one obtained from the web covering celebrities. We developed classification models that rely on a combination of lexical, syntactic, and semantic information, as well features representing text readability properties. Our best per-forming models achieved accuracies that are com-parable to human ability to spot fake content.

This paper is based on the crowd sourcing dataset and the web covering dataset. These are the static datasets. Through these we can only test the data which is present in the predefined training data sets. The paper gives the appropriate result for the test data that is present in the training datasets. Thus, the future scope of the paper is connecting this methodology to the internet news which gives results even for the test data that is not present in the training data sets. We can even change to some other better classifier to classify the data other than naïve bayes and logistic regression.

REFERENCES

- [1] Majed Alrubaian, Muhammad Al-Qurishi, A Credibility Analysis System for Assessing Information on Twitter, IEEE Transactions on Dependable and Secure Computing, 1-14. DOI : <http://dx.doi.org/10.1109/TDSC.2016.2602338>
- [2] Manish Gupta, Peixiang Zhao, Jiawei Han, 2012. Evaluating Event Credibility on Twitter, Proceedings of the 2012 SIAM International Conference on Data Mining, 153-164, DOI: <http://epubs.siam.org/doi/abs/10.1137/1.9781611972825.14>
- [3] Krzysztof Lorek, Jacek Suehiro-Wiciński, Michał Jankowski-Lorek, Amit Gupta, Automated credibility assessment on twitter, Computer Science, 2015, Vol.16(2), 157-168, DOI: <http://dx.doi.org/10.7494/csci.2015.16.2.157>
- [4] Ballouli, Rim El, Wassim El-Hajj, Ahmad Ghandour, Shady Elbassuoni, Hazem M. Hajj and Khaled Bashir Shaban. 2017. "CAT: Credibility Analysis of Arabic Content on Twitter." WANLP@EACL (2017).
- [5] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.
- [6] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Association for Computational Linguistics, 2012, pp. 171–175.
- [7] Granik, M., & Mesyura, V. 2017. Fake news detection using naive Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 900-903.

ACKNOWLEDGEMENT

We have great pleasure in expressing our gratitude to Sri K.V.V.Satyanarayana Raju, Founder & Chairman, Chaitanya Group of Institutions, Sri K. Sasi Kiran Varma, Vice Chairman, GIET Group of Institutions, Smt. Lakshmi Raju Executive Director, GIET, for their kind support in providing us an opportunity to do research in this college.