



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Fake News Detection Using Machine Learning

Arshad Gilani, Manish Banajwade, Akshay Lonkar, Rakesh Kumawat

Dept. of Computer Engineering, Keystone School of Engineering, Pune, Maharashtra, India

ABSTRACT: Fake news detection is an interesting topic for computer scientists and social science. The recent growth of the online social media fake news has great impact to the society. There is a huge information from disparate sources among various users around the world. Social media platforms like Facebook, WhatsApp and Twitter are one of the most popular applications that are able to deliver appealing data in timely manner. Developing a technique that can detect fake news from these platforms is becoming a necessary and challenging task. This project proposes a machine learning method which can identify the credibility of an article that will be extracted from the Uniform Resource Locator (URL) entered by the user on the front end of a website. The project uses the two widely used machine learning methods: Random Forest (random tree), Random Forest (decision tree), Decision Tree and to give a response telling the user about the credibility of that news. Our initial definition of reliable and unreliable will rely on the human-curated data <http://opensources.co>. OpenSources.co has a list of about 20 credible news websites and a list of over 700 fake news websites.

KEYWORDS: Machine Learning, Artificial Intelligence, Fake news Detection, Dataset, Training

I. INTRODUCTION

An increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations. The explanations for this alteration in consumption behaviors are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism, like newspapers or television; and (ii) it's easier to further share, discuss, and discuss the news with friends or other readers on social media. For instance, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media. It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. It had been estimated that over 1 million tweets are associated with fake news "Pizzagate" by the top of the presidential election. Given the prevalence of this new phenomenon, Fake news" was even named the word of the year by the Macquarie dictionary in 2016. The extensive spread of faux news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem for instance; it's evident that the most popular fake news was even more outspread on Facebook than the most accepted genuine mainstream news during the U.S. 2016 presidential election. Second, fake news intentionally persuades consumers to simply accept biased or false beliefs. Fake news is typically manipulated by propagandists to convey political messages or influence for instance, some report shows that Russia has created fake accounts and social bots to spread false stories. Third, fake news changes the way people interpret and answer real news, for instance, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what's true from what's not. To assist mitigate the negative effects caused by fake news (both to profit the general public and therefore the news ecosystem). It's crucial that we build up methods to automatically detect fake news broadcast on social media.

II. PROBLEM DEFINITION

Fake news is written and published usually with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically, often using sensationalist, dishonest, or outright fabricated headlines to increase reader-ship. In the past presidential election, the American people were overwhelmed with the proliferation of "fake news" articles that altered the narrative (and perhaps the results) of the election. The articles and social media posts featured bombastic headlines and made outrageous claims regarding the candidates. Indians have become distrustful of mainstream news outlets and are spreading information from alternative sources without verifying it. While doing so, they

believe themselves to be promoting the real story. India has recently witnessed several shocking incidents of mob lynching caused by the spread of rumors over What- sApp. Fake news related to cow slaughter and child abduction is spread via the social media platforms resulting in the mob agitation. Instigated by the messages from an unknown source people opt to become the vigilantes. This scenario has resulted in the defiance of the law, authority and killing of the many innocent citizens without any fair judicial intervention.

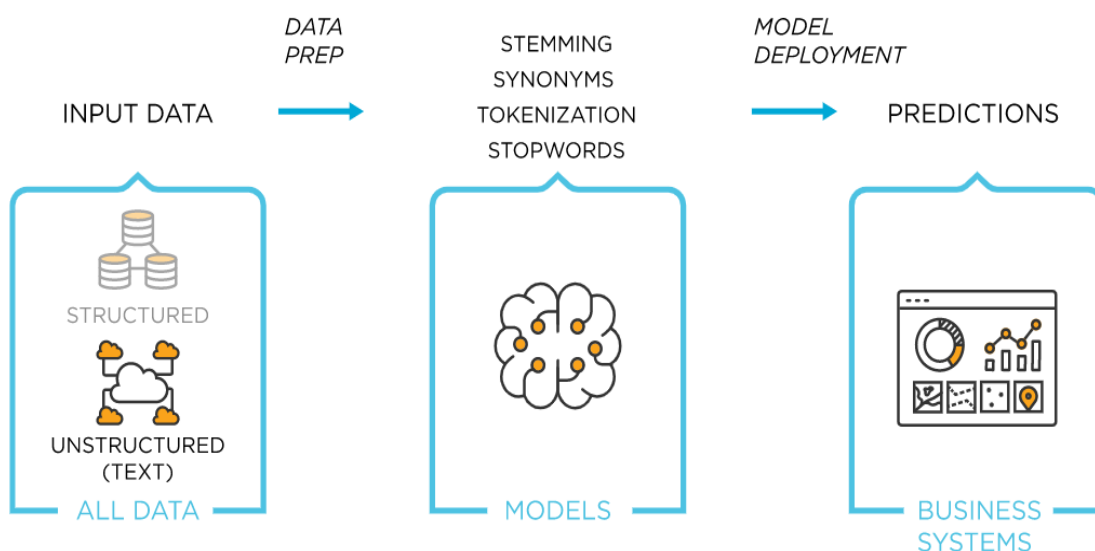
III. METHODOLOGY

This paper explains the system which is developed in three parts. The first part is static which works on machine learning classifier. We studied and trained the model with 4 different classifiers and chose the best classifier for final execution. The second part is dynamic which takes the keyword/text from user and searches online for the truth probability of the news. The third part provides the authenticity of the URL input by user. In this paper, we have used Python and its Sci-kit libraries. Python has a huge set of libraries and extensions, which can be easily used in Machine Learning. Sci-Kit Learn library is the best source for machine learning algorithms where nearly all types of machine learning algorithms are readily available for Python, thus easy and quick evaluation of ML algorithms is possible. We have used Django for the web based deployment of the model, provides client side implementation using HTML, CSS and Javascript. We have also used Beautiful Soup (bs4), requests for online scrapping. System Design-Figure 1: System Design System Architecture- Static Search-The architecture of Static part of fake news detection system is quite simple and is done keeping in mind the basic machine learning process flow. The system design is shown below and self- explanatory. The main processes in the design are- Figure 2: System Architecture Dynamic Search-The second search field of the site asks for specific keywords to be searched on the net upon which it provides a suitable output for the percentage probability of that term actually being present in an article or a similar article with those keyword references in it. URL Search- The third search field of the site accepts a specific website domain name upon which the implementation looks for the site in our true sites database or the blacklisted sites database. The true sites database holds the domain names which regularly provide proper and authentic news and vice versa. If the site isnt found in either of the databases then the implementation doesnt classify the domain it simply states that the news aggregator does not exist

IV. STEPS OF IMPLEMENTATION

TEXT PREPARATION

Social media data is highly unstructured – majority of them are informal communication with typos, slangs and bad-grammar etc. To achieve better insights, it is necessary to clean the data before it can be used for predictive modeling. For this purpose, basic pre-processing was done on the News training data. This step was comprised of



1. **Conversion to Lower case:** First step was to transform the text into lower case, just to avoid multiple copies of the same words. For e.g. while finding the word count, “Response” and “response” is taken as different words.
2. **Removal of Punctuations:** Punctuations does not have much significance while treating the text data. Therefore, removing them helps to reduce the size of overall text
3. **Stop-words removal:** Stop-words are the most commonly occurring used words in a corpus. These are for e.g. a, the, of, on, at etc. They usually define the structure of a text and not the context. If treated as feature, they would result in poor performance. Therefore, Stop-words were removed from the training data as the part of text cleaning process.
4. **Tokenization:** It refers to dividing the text into a sequence of words or group of words like bigram, trigram etc. Tokenization was done so that frequency- based vectors values could be obtained for these tokens.
5. **Lemmatization:** It converts the words into its word root. With the help of a vocabulary, it does morphological analysis to pick up the root word. In this work, Lemmatization was performed to improve the values of frequency-based vectors.

Text pre-processing was an essential step before the data was ready for analysis. A noise free corpus has a reduced size of the sample space for features thereby resulting in increased accuracy.

SYSTEM ARCHITECTURE

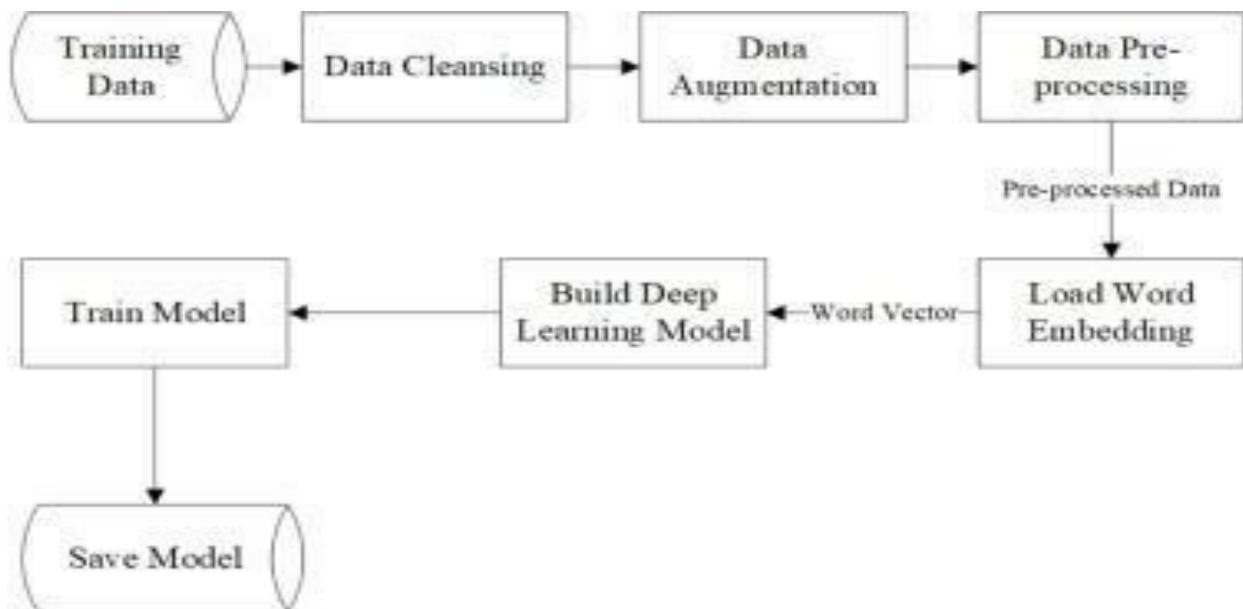


Fig.1 System Architecture

Data Flow Diagram

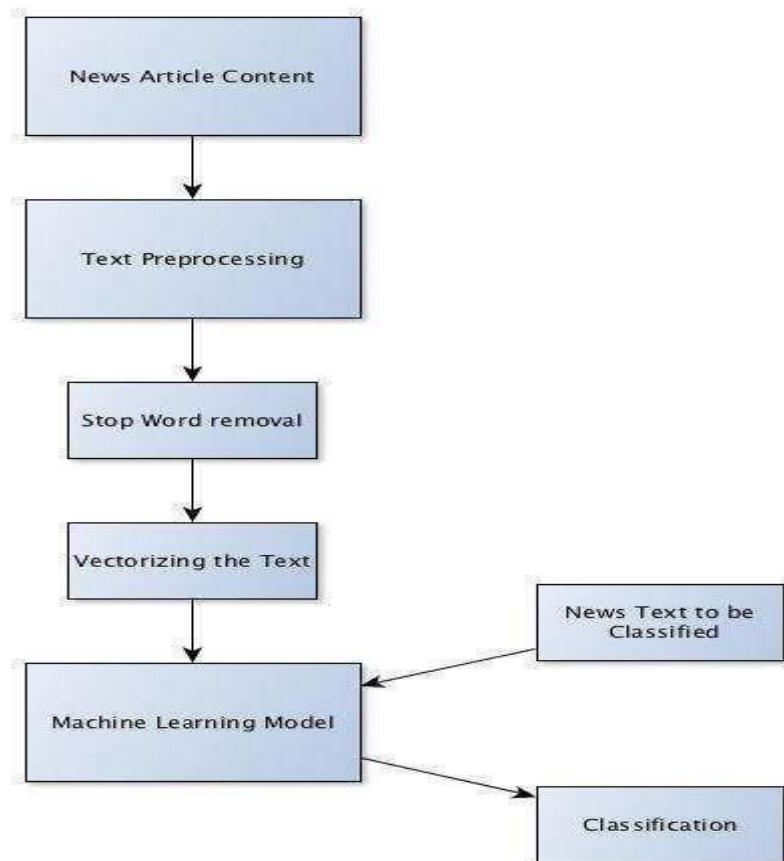


Fig.2 Data Flow Diagram

SOFTWARE USED

- Programming language used: Python
- Framework used: Django
- Libraries used: OpenCV, Numpy, Sklearn, pyttsx3
- Code Style Format: PEP8

V. CONCLUSION

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has strong negative impacts on individual users and broader society. This fake news is created by someone intentional to create wanted violence, anger in our society. Most of the times, the younger generations face the lot of effects mentally. To avoid this situation, we have created this system called fake news detector. Fake news detection system will differentiate between fake and real news from the social media. We have explored machine learning techniques to detect the news, real or fake. In our project we have characterized two parts i.e., parsing of data and classification. For parsing of data, we have explored many libraries, tools, but the simplest and easiest way was through using python libraries i.e., request library

and HTML5lib parser library. By using them we were able to extract the data and store in structured form. included with this system to make it a more robust and error free computer-based automatic system for surveillance.

REFERENCES

1. Fang Jin et al. "Epidemiological Modeling of News and Rumors on Twitter". In: Proceedings of the 7th Workshop on Social Network Mining and Analysis. SNAKDD '13. Chicago, Illinois:
2. ACM, 2013, 8:1–8:9. ISBN : 978-1-4503-2330-7. DOI : 10.1145/2501025.2501027. URL : <http://doi.acm.org/10.1145/2501025.2501027>.
3. Yang P. Liu and Yi-fang Brook Wu. "Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks". In: AAAI. 2018.
4. Yang Liu and Yi-Fang Brook Wu. "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks". In: Thirty-Second AAAI Conference on Artificial Intelligence. 2018
5. V. Kecman, *Support Vector Machines-An Introduction in "Support Vector Machines: Theory and Applications"*, Springer, New York City, NY, USA, 2005.
6. S. Akhtar, F. Hussain, F. R. Raja et al., "Improving mispronunciation detection of arabic words for non-native learners using deep convolutional neural network features," *Electronics*, vol. 9, no. 6, 2020.View at: [Publisher Site](#) | [Google Scholar](#)
7. D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Information Fusion*, vol. 6, no. 1, pp. 63–81, 2005.View at: [Publisher Site](#) | [Google Scholar](#)
8. B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests," *Statistics and Computing*, vol. 27, no. 3, pp. 659–678, 2017.View at: [Publisher Site](#) | [Google Scholar](#)
9. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Springer, Berlin, Germany, 1984.
10. R. E. Schapire, "A brief introduction to boosting," *IJCAI*, vol. 99, pp. 1401–1406, 1999.View at: [Google Scholar](#)
11. E. M. Dos Santos, R. Sabourin, and P. Maupin, "Overfitting cautious selection of classifier ensembles with genetic algorithms," *Information Fusion*, vol. 10, no. 2, pp. 150–162, 2009.View at: [Publisher Site](#) | [Google Scholar](#)
12. T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco, CA, USA, 2016.View at: [Google Scholar](#)
13. T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.View at: [Publisher Site](#) | [Google Scholar](#)



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details