

A Review On Detection of Fake News Using Various Techniques

Mauik Panchal¹, Prof. Rutika Ghariya²

M.E. Computer Engineering Department (Cyber Security), GTU-Graduate School of Engineering and Technology, Ahmedabad, India
Assistant Professor, GTU – GSET, GTU-Graduate School of Engineering and Technology, Ahmedabad, India

Received Date: 22 May 2021

Revised Date: 26 June 2021

Accepted Date: 02 July 2021

Abstract — For past many years, since the rise of social media, fake news has become a genuine society problem, in some occasion this news spreads faster than the true information. Social media interaction especially the news spreading around the network is a great source of information nowadays. From one's perspective, its negligible exertion, straightforward access, and quick dispersing of information that lead people to look out and trust news from internet-based life. Twitter being a standout amongst the most well-known ongoing news sources additionally ends up standout amongst the most dominant news radiating mediums. It is known to cause extensive harm by spreading bits of gossip previously. Online clients are normally vulnerable and will, in general, perceive all that they run over web-based networking media as reliable. This paper proposes a model for recognizing fake news messages from all news medium, by figuring out how to anticipate precision appraisals, in view of computerizing fake news identification in different mediums. Afterwards, we performed a comparison between well-known Machine Learning algorithms, like Support Vector Machine, Naïve Bayes Method, Logistic Regression and Recurrent Neural Network models, separately to demonstrate the efficiency of the classification performance on the dataset. According to all those analyses it is possible to wipe out this widespread of such fake news using machine learning and serve people correct news which is essential to them.

Keywords — Machine Learning, Random Forest, NLP techniques, Support Vector Machine, Fake News Detection

I. INTRODUCTION

Fake news is something which is very common hazard to people these days. Due to misleading news, there can be serious mishaps. People these days, with raising use of social media, trust the internet shared or social media shared news blindly rather than actually evaluating it against the real news. It is very essential to differentiate between the real news and personal point of view shared on the social media. This fake news which are trusted by normal people will have great impact and can also lead to serious calamities which can also be life threatening at times. In recent scenario, when the world is facing covid-19 there is a fake news wave going on social media, through Instagram WhatsApp etc. people who are unaware about the actual facts blindly believe in that forwarded news

which leads to trust issues, panic, also depression in some cases. Fake news about government guidelines, fake news about some riots, some calamities are most popular scenarios of hazards to mankind. An intense inspection the present tweets demonstrates that false news spreads frequently through human than a genuine news does. Lie gets travelled around us quicker, and more extensively than reality in all spheres of information, and the effects were more dangerous and horrifying. There are several kinds of tweets like issues on a government, trending topics around the world, mental abuse, urban legends, occasions in calamities etc. There are many problems arising due to translation of real news into news which has provoking statements to common people also which are misleading which creates trust issue towards one another, also government. In recent times, we can see there are lot many news about covid 19, farmer's protest, about different government acts etc. Some people take benefit of those fake news for their personal benefits. There are many new products launched in market for which again some new fake promotions are done. For example, there are many medicines to boost immunity or some different hand washes available in market which will fight covid. Fake news is actually increasing people's fear. Which is in turn helping many smart minds to get business. To evaluate such news and put it in real manner it is very much essential to detect what is fake and what is real. There might be such cases that the real news is presented in such a way that it provokes mindset. So it is essential to understand the essence of the real news and putting in a language which is lucid to make common people understand about what is really happening. This not only helps to build a trustworthy environment but will also boost confidence on many ends such as people will be more inclined. The current paper evaluates the news in two parts authenticity and intent. Authenticity means that fake news content false information that can be verified as such, which means that conspiracy theory is not included in fake news as there are difficult to be proven true or false in most cases. The second part, intent, means that the false information has been written with the goal of misleading the reader.

II. BACKGROUND AND MOTIVATION

Fake news spreads from sources to consumers through a complex ecosystem of websites, social media, and bots. Features that make social media engaging, including the



ease of sharing and rewiring social connections, facilitate their manipulation by highly active and partisan individuals (and bots) that become powerful sources of misinformation.

Even if individuals prefer to share high-quality information, limited individual attention and information overload prevent social networks from discriminating between messages on the basis of quality at the system level, allowing low-quality information to spread as virally as high-quality information. This helps explain higher exposure to fake news online.

Most people who share fake news, whether it gains popularity or not, share lots of news in general. Volume of political activity is by far the strongest predictor of whether an individual will share a fake news story. The fact that misinformation is mixed with other content and that many stories get little attention from people means that traditional measures of quality cannot distinguish misinformation from truth.

Most of us do not witness news events first hand, nor do we have direct exposure to the workings of politics. Instead, we rely on accounts of others; much of what we claim to know is actually distributed knowledge that has been acquired, stored, and transmitted by others. Likewise, much of our decision-making stems not from individual rationality but from shared group-level narratives. As a result, our receptivity to information and misinformation depends less than we might expect on rational evaluation and more on the heuristics and social processes we describe below. First, source credibility profoundly affects the social interpretation of information. Individuals trust information coming from well-known or familiar sources and from sources that align with their worldview. Second, humans are biased information-seekers: we prefer to receive information that confirms our existing views. When a source is perceived as credible or the information confirms prior views. And when the information is unfamiliar or comes from an opposition source, it may be ignored. As a result, correcting misinformation does not necessarily change people’s beliefs. In fact, presenting people with challenging information can even backfire, further entrenching people in their initial beliefs. However, even when an individual believes the correction, the misinformation may persist. An important implication of this point is that any repetition of misinformation, even in the context of refuting it, can be harmful. Encouraging communication with people who are dissimilar might be an effective way to reduce polarization.

III. LITURATURE REVIEW

- In [1] discuss about fake news detection and ways to apply them on various Social media sites using naïve Bayes classifier. The data sources for news article are Facebook, twitter etc. The accuracy obtained is quite low as these site’s information is not 100% credible
- In [2] evaluated different machine learning algorithms and analyzed the prediction

percentage. The accuracy of different predictive models which included bounded decision trees, gradient boosting, and support vector machine were tabulated. The models are evaluated on the basis of probability threshold which aren’t most reliable.

- In [3] detecting fake news through various machine learning models. The given machine learning models implemented are naïve Bayes classifier and support vector machine. No specific accuracy was recorded as only the models were discussed.
- In [4] to detect fake news on social media, presents a data mining perspective which includes fake news characterization on psychology and social theories. This article discusses two major factors responsible for widespread acceptance of fake news by the user which are Naive Realism and Confirmation Bias.
- In [5] approach on the fake news was detected using Ngram analysis has described how the fake news has rendered in to social media and that false news has been wide spread between youngsters, it grabbed the attention of the youngsters to believe only the social media news not the printed news. Using some feature extraction techniques, and investigated on different feature extraction and machine learning techniques.

IV. COMPARATIVE ANALYSIS

TABLE I
ANNOTATION EXAMPLES OF
TEXT CATEGORIES RELATED
TO FAKE NEWS DETECTION

Literature	Dataset	Models	Accuracy
Fake News Detection Using Machine Learning approaches: A systematic Review.[1]	Facebook	.Decision trees, Random Forest, SVM	70%
Behind the cues: A benchmarking study for fake news detection .[3]	Open sources, George McIntire dataset	SVC, RN, AB, MLP	86%
Fake News Detection Using Machine Learning. [6]	Weibo	KNN, NB, SVM, XGB	83.1%
Fake News Detection. [7]	Social Media	SVM	72%

V. METHODOLOGY

A. User:

The person who is using this Service.

B. Dataset:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

C. Data Processing:

Data in its raw form is not useful to any organization. Data processing is the method of collecting raw data and translating it into usable information. It is usually performed in a step-by-step process by a team of data scientists and data engineers in an organization. The raw data is collected, filtered, sorted, processed, analyzed, stored and then pre-sented in a readable format. Data processing is, generally, "the collection and manipulation of items of data to produce meaningful information." In this sense it can be considered a subset of information processing, "the change (processing) of information in any manner detectable by an observer."

D. Types of data processing include:

a) Manual data processing: Manual data processing refers to data processing that requires humans to manage and process the data throughout its existence. Manual data processing utilizes non-technological tools, which include paper, writing utensils and physical filing cabinets.

b) Automatic data processing: An interacting assembly of procedures, processes, methods, personnel, and equipment to perform automatically a series of data processing operations on data.

Note: The data processing operations may result in a change in the semantic content of the data. Data processing by means of one or more devices that use common storage for all or part of a computer program, and also for all or part of the data necessary for execution of the program; that execute user-written or user-designated programs; that perform user-designated symbol manipulation, such as arithmetic operations, logic operation, or character-string manipulations; and that can execute programs that modify themselves during their execution. Note: Automatic data processing may be

performed by a stand-alone unit or by several connected units.

c) Electronic data processing: Electronic data processing is a generic term that signifies the gathering, intelligent analysis, and manipulation of data that is put to work. Warehouse stock monitoring and logistics system is one of the most common examples of EDP in the modern age. It is efficient. Summary documents and related materials such as invoices, reports, and statements can be automatically and quickly generated via EDP. Economic. Once an Electronic Data Processing system is created and implemented, over time it reduced the costs of managing data by a significant margin.

d) Word to vector conversion: Converting words to vectors, or word vectorization, is a natural language processing (NLP) process. The process uses language models to map words into vector space. A vector space represents each word by a vector of real numbers. It also allows words with similar meanings have similar representations. This module requires a dataset that contains a column of text. Preprocessed text is better.

- Add the Convert Word to Vector module to your pipeline.
- As input for the module, provide a dataset that contains one or more text columns.
- For Target column, choose only one column that contains text to process.
- Because this module creates a vocabulary from text, the content of columns differs, which leads to different vocabulary contents. That's why the module accepts only one target column.
- For Word2Vec strategy, choose from GloVe pretrained English Model, Gensim Word2Vec, and Gensim FastText.

• If Word2Vec strategy is Gensim Word2Vec or Gensim FastText:

1. For Word2Vec Training Algorithm, choose from Skipgram and CBOW. The difference is introduced in the original paper(PDF).
2. The default method is Skipgram.
3. For Length of word embedding, specify the dimensionality of the word vectors. This setting corresponds to the size parameter in Gensim.
4. The default embedding size is 100.
5. For Context window size, specify the maximum distance between the word being predicted and the current word. This setting corresponds to the window parameter in Gensim.
6. The default window size is 5.
7. For Number of epochs, specify the number of epochs (iterations) over the corpus. Corresponds to the iter

parameter in Gensim. The default epoch number is 5.

8. For Maximum vocabulary size, specify the maximum number of the words in the generated vocabulary. If there are more unique words than the max size, prune the infrequent ones. The default vocabulary size is 10,000.
9. For Minimum word count, provide a minimum word count. The module will ignore all words that have a frequency lower than this value. The default value is 5.

Submit the pipeline. Examples The module has one output: Vocabulary with embeddings: Contains the generated vocabulary, together with each word's embedding. One dimension occupies one column. The following example shows how the Convert Word to Vector module works. It uses Convert Word to Vector with default settings to the preprocessed Wikipedia SP 500 Dataset.

VI. CONCLUSIONS

In this work, we propose a various machine learning methods. We used the word to vector mechanism to convert the news article into a news vector which contains the important features which are used to determine the nature of the content. With the help of Bag of Words (Bow) vector, extracting features from text. Bow identify the Vocabulary of known words & measure of the presence of known words to collect all the unique words in all the article content. After generating the word methodology, with the help of Machine Learning Algorithm We classified the vector to check whether it is fake news or real news.

VII. REFERENCES

- [1] Manzoor, S. I., Singla, J., Nikita., Fake News Detection Using Machine Learning approaches: A systematic Review, 3rd International Conference on Trends in Electronics and Informatics(ICOEI), (2019) 230–234. IEEE.
- [2] Han, W., Mehta V., Fake news detection in social networks using machine learning and deep learning: Performance evaluation, IEEE International Conference on Industrial Internet (ICII), (2019) 375-380. IEEE.
- [3] Gravanis, G., Vakali, A., Diamantaras, K., Karadais, P., Behind the cues: A benchmarking study for fake news detection. Expert Systems with Applications, 128 (2019) 201–213.
- [4] Bali, A. P. S., Fernandes, M., Choubey, S., Goel, M., Comparative performance of machine learning algorithms for fake news detection, In Communications in Computer and Information Science, (2019) 420–430. Singapore: Springer Singapore.
- [5] Lakshmanarao, A., Swathi, Y., Srinivasa Ravi Kidetection system us-ing machine learning. International Journal ran, T., An efficient fake news of Innovative Technology and Exploring Engineering, 8(10) (2019)3125–3129.
- [6] Ibrishimova, M. D., Li, K. F., A machine learning approach to fake news detection using knowledge verification and natural language processing, In Advances in Intelligent Networking and Collaborative Systems, (2020) 223–234. Cham: Springer International Publishing.

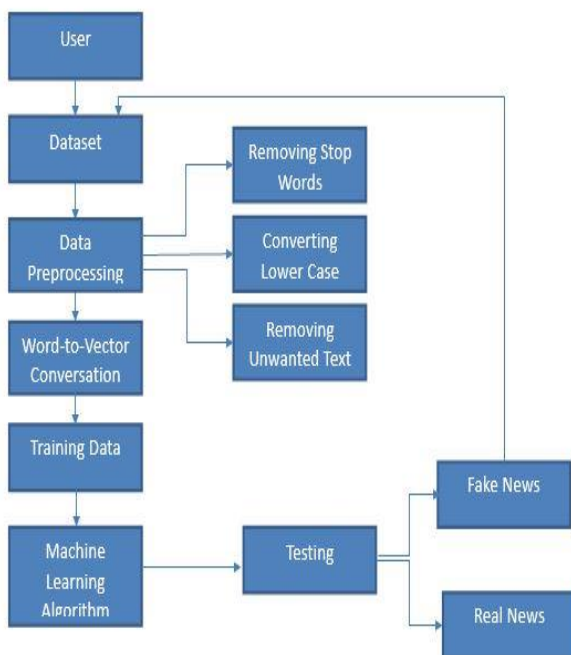


Fig. 1. Fake News Detection Model