

FAKE NEWS DETECTION BY USING NATURAL LANGUAGE PROCESSING

Mr. Esham Rishal

B. E. Computer Science and Engineering
Hindusthan College of Engineering and Technology
Coimbatore, India
20104023@hicet.ac.in

Mr. Sankaranarayanan K

B. E. Computer Science and Engineering
Hindusthan College of Engineering and Technology
Coimbatore, India
20104053@hicet.ac.in

Mr. Shahiz Shaheer

B. E. Computer Science and Engineering
Hindusthan College of Engineering and Technology
Coimbatore, India
20104055@hicet.ac.in

Ms. N. PRIYA.

B. E. Computer Science and Engineering
Hindusthan College of Engineering and Technology
Coimbatore, India
npriya.cse@hicet.ac.in

Mr. Devakrishna P P

B. E. Computer Science and Engineering
Hindusthan College of Engineering and Technology
Coimbatore, India
20104019@hicet.ac.in

Abstract

In recent years, due to the booming development of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely. This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles, creators and subjects from online social networks and evaluating the corresponding performance. Information preciseness on Internet, especially on social media, is an increasingly important concern, but web-scale data hampers, ability to identify, evaluate and correct such data, or so called "fake news," present in these platforms. In this paper, we propose a method for "fake news" detection and ways to apply it on Facebook, one of the most popular online social media platforms. This method uses NaiveBayes classification model to predict whether a post on Facebook will be labeled as real or fake. The results may be improved by applying several techniques that are discussed in the paper. Received results suggest, that fake news detection problem can be addressed with machine learning method.

I. 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 see's 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 .

II . Literature Review

I. Niall J. Conroy et all in Automatic Deception Detection: Methods for Finding Fake News 2015, made this research surveys the current state-of-the art technologies that are instrumental in the adoption and development of fake news detection. “Fake news detection” is defined as the task of categorizing news along a continuum of veracity, with an associated measure of certainty. Veracity is compromised by the occurrence of intentional deceptions

II. Yimin Chen et all in Misleading online content: Recognizing clickbaits as false news 2015, Tabloid journalism is often criticized for its propensity for exaggeration, sensationalization, scare-mongering, and otherwise producing misleading and low quality news. As the news has moved online, a new form of tabloidization has emerged: „clickbaiting.“ „Clickbait“ refers to “content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page” [„clickbait,” n.d.] and has been implicated in the rapid spread of rumor and misinformation online.

III. Maryam M Najafabadi et all in Deep learning applications and challenges in big data analytics 2015, Big Data Analytics and Deep Learning are two high- focus of data science. Big Data has become important as many organizations both public and private have been collecting massive amounts of domain-specific information, which can contain useful information about problems such as national intelligence, cyber security, fraud detection, marketing, and medical informatics. Companies such as Google and Microsoft are analyzing large volumes of data for business analysis and decisions, impacting existing and future technology.

IV. Eugenio Tacchini et all in Some like it hoax: Automated fake news detection in social networks 2017, In recent years, the reliability of information on the Internet has emerged as a crucial issue of modern society. Social network sites (SNSs) have revolutionized the way in which information is spread by allowing users to freely share content. As a consequence, SNSs are also increasingly used as vectors for the diffusion of misinformation and hoaxes. The amount of disseminated information and the rapidity of its diffusion make it practically impossible to assess reliability in a timely manner, highlighting the need for automatic hoax detection systems

V. Chengcheng Shao et all in The spread of fake news by social bots 2017, The massive spread of fake news has been identified as a major global risk and has been alleged to influence elections and threaten democracies. Communication, cognitive, social, and computer scientists are engaged in efforts to study the complex causes for the viral diffusion of digital misinformation and to develop solutions, while search and social media platforms are beginning to deploy countermeasures.

VI. Shivam B. Parikh et all in Media-Rich Fake News Detection: A Survey 2018, Fake News has been around for decades and with the advent of social media and modern day journalism at its peak, detection of media-rich fake news has been a popular topic in the research community. Given the challenges associated with detecting fake news research problem, researchers around the globe are trying to understand the basic characteristics of the problem statement. This paper aims to present an insight on characterization of news story in the modern diaspora combined with the differential content types of news story and its impact on readers

VII. Della Vedova et all in Automatic Online Fake News Detection Combining Content and Social Signals 2018, The proliferation and rapid diffusion of fake news on the Internet highlight the need of automatic hoax detection systems. In the context of social networks, machine learning (ML) methods can be used for this purpose.

III. Problem and Existing System

- The spread of fake news, which includes false or misleading information presented as news, poses significant risks such as misinforming the public, influencing political outcomes, causing economic disruptions, and fueling social unrest. Identifying and mitigating fake news is challenging due to the vast amount of information online and sophisticated methods used to create believable false content.
- **EXISTING SYSTEM**
 - **Rule-based System:**
 - Manual Checking: Human fact-checkers verify news, but it's labor-intensive and not scalable.
 - Keyword Filtering: Flags potential fake news using predefined keywords, but often lacks accuracy.
 - **Traditional Machine Learning:**
 - Feature Engineering: Uses crafted features like word patterns and source credibility.
 - Supervised Models: Classifiers such as Naive Bayes, SVM, and Random Forests trained on labeled datasets.
 - **Deep Learning techniques:**
 - Neural Networks: CNNs and RNNs automatically learn features from text.
 - Transformers: Models like BERT and GPT excel in understanding context and nuances.

IV. System Architecture

1. Data Collection and Preprocessing

- **Data Sources:** Collect data from various sources such as news websites, social media platforms, and RSS feeds.
- **Data Preprocessing:** Clean and preprocess the text data by removing HTML tags, special character and stop words, and performing stemming/lemmatization.

2. Feature Extraction

- **Textual Features:** Extract features from the text such as:
 - **N-grams:** Unigrams, bigrams, trigrams
 - **TF-IDF (Term Frequency-Inverse Document Frequency)**
 - **Word Embeddings:** Using pre-trained models like Word2Vec, GloVe, or contextual embeddings like BERT.
- **Metadata Features:** Source credibility, publication date, author information, and user engagement metrics.
- **Content Features:** Linguistic and stylistic features, such as sentiment analysis and readability scores.

3. Model Building

- **Training and Testing Datasets:** Split the preprocessed data into training and testing sets.
- **Model Selection:** Choose appropriate machine learning or deep learning models..
- **Training:** Train the models on the training dataset.
- **Validation:** Validate model performance using cross-validation techniques.

4. Model Evaluation and Optimization

- **Evaluation Metrics:** Use metrics like accuracy, precision, recall, F1-score, and ROC-AUC to evaluate model performance.
- **Hyperparameter Tuning:** Optimize model hyperparameters to improve performance.

- **Ensemble Methods:** Combine multiple models to enhance robustness and accuracy.

5. Deployment

- **Model Deployment:** Deploy the trained model using a scalable infrastructure such as cloud services (AWS, Azure, GCP).

6. User Interface

- **Dashboard:** Develop a dashboard for users to input news articles and receive fake news detection results.
- **Visualization:** Provide visualizations of data insights and detection results.

V. Architecture Diagram

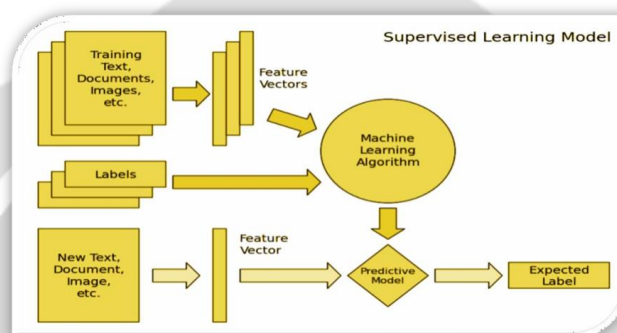


Fig. 2. Architecture Diagram of system.

VI. Implementation and Deployment

1. Importing packages such as pandas, numpy, csv, and importing algorithms from sklearn as we are using the Python Library.
2. Reading the CSV file and displaying the data set.
3. Remove the erroneous null and nan values.
4. Calculating the length of news so as to check the length of fake and real news.
5. Data Preprocessing : NLTK tokenization Tokenizing and normalizing the text Lemmatization and Vectorization
6. Feature Extraction
7. Training a model to detect fake
8. Build and evaluate
9. Evaluate Accuracies

VI. Results and Discussion

- 1) Algorithm's accuracy depends on the type and size of your dataset. More the data, more chances of getting correct accuracy. Machine learning depends on the variations and relations. Understanding what is predictable is as important as trying to predict it. While making algorithm choice, speed should be a consideration factor. The project highlights the results of the Application and the snapshots for each of the activities are shown along with the discussion of each activity describing its working. Each snapshots describes every single step of making Fake news detect

ID	title	author	text	label
1	House Dem Aid: We Didn't Even See Comey's Letter Until Jason			1
2	Why the Truth Consumption: Why the Truth Might Get You Freed October 25, 2016			0
3	15 Civilians Killed in Single US Airstrike Have Been Identified The			1
4	Iranian woman Howard Portnoy			1
5	Jackie Mason: Daniel Nussbaum			0
6	Life of Luan			1
7	France chose an idealistic, traditional candidate in Sunday's			0
8	Donald Trump is scheduled to make a highly anticipated visit to an			0
9	A Back-Channe Megan Tachey and A week before Michael T. Flynn resigned as national security adviser, a seale			0
10	Obama's Organizing for Action, the activist group that morphed from Barack Obama's			0
11	The BBC produced spoof on the Alkaalil Housewives TV programmes, whi			0
12	Clinton Campaign Demands FBI Affair Trump's Russia Ties			1
13	Yes, There Are Paid Government Trolls On Social Media, Blogs, Forums And Websites			1
14	Do you want solid pro			0
15	Guillermo Barros Schelotto was not the first Argentine player to set foot on a			0
16	The scandal engulfing Wells Fargo toppled its chairman and chief executive c			0
17	A Caddo Nation tribal leader has just been freed after spending two days			1
18	FBI Closes in On Hillary's Potted on Home & Headlines & World News &			1
19	Wednesday after Donald Trump's press conference at Trump Tower in N			0
20	Email			1

Fig. 1. training dataset

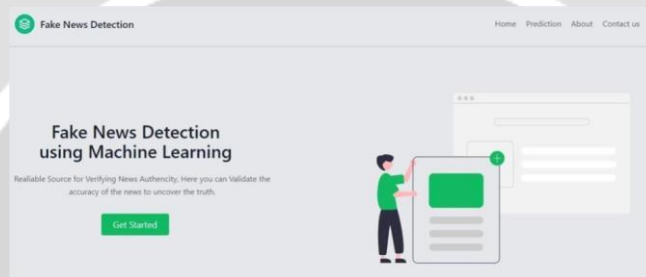


Fig. 3. Webpage

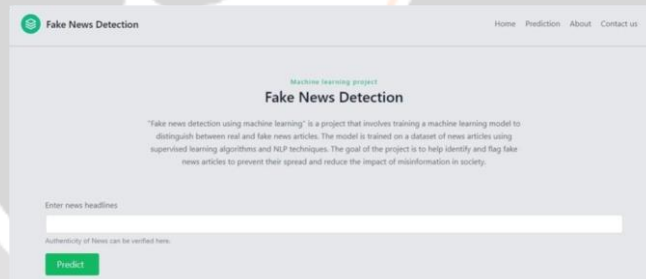


Fig. 4. Entering the article

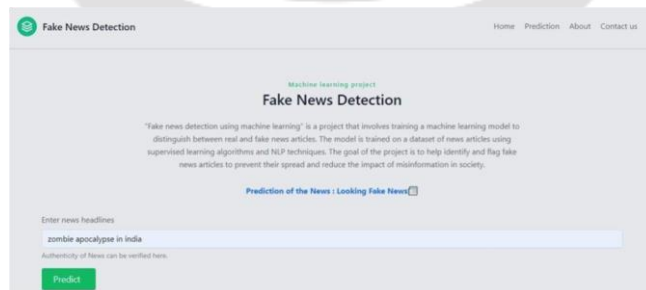


Fig.5. Output of entered article

VII. CONCLUSIONS

Many people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has negative impacts on individual people and society. In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. This model takes news events as an input and based on twitter reviews and classification algorithms it predicts the percentage of news being fake or real.

VIII. REFERENCES

- [1] Chen, Y., Conroy, N. J., & Rubin, V. L. (2015, November). Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection (pp. 15-19). ACM.
- [2] Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 82). American Society for Information Science..
- [3] Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L. (2018, May). Automatic Online Fake News Detection Combining Content and Social Signals. In 2018 22nd Conference of Open Innovations Association (FRUCT) (pp. 272-279). IEEE.
- [4] Haiden, L., & Althuis, J. (2018). The Definitional Challenges of Fake News. .
- [5] Helmstetter, S., & Paulheim, H. (2018, August). Weakly supervised learning for fake news detection on Twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 274-277). IEEE.
- [6] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1.
- [7] Parikh, S. B., & Atrey, P. K. (2018, April). Media-Rich Fake News Detection: A Survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 436-441). IEEE.
- [8] Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, 96-104.
- [9] Stahl, K. (2018). Fake News Detection in Social Media.