



IJIRCCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 12, December 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Fake News Detection Using Machine Learning

Rutuja Gandhe, Anamika Choudhary, Aditya Dagda, Prof. Vajid Khan sir

Department of Computer Engineering, JSCOE, Hadapsar, Pune, India

ABSTRACT: Most smartphone users prefer to read news on social media over the internet. The News website publishes news and provides a source of certification. The question is how to certify news and articles broadcast on social media such as WhatsApp groups, Facebook pages, Twitter and other micro-blogs and social networking sites. Believing in the rumors and pretending that it is a news is harmful to society. The need for hours is to stop rumors, especially in developing countries like India, and focus on the correct certified news articles of . This paper presents a model and methodology for fake news detection. It uses machine learning and natural language processing to aggregate messages and later uses a support vector machine to try to determine if the message is real or fake. The results of the proposed model will be compared to the existing model. The proposed model works well and defines the accuracy of news in the results up to 93.6% of the accuracy.

I. INTRODUCTION

As 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.

1] The explanations for this alteration in consumption behaviour 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;

(ii) it's easier to further share, discuss , and discuss the news with friends or other readers on social media. But out of them 65% are fake .we should clarify the news before forwarding.

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 fake 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. LITERATURE SURVEY

●MykhailoGranik et. al. in their paper shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

●Himank Gupta et. al. gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (Bot-Maker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150,000 spam tweets and 250,000 non- spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of-Words model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately 18%.

III. METHODOLOGY

Fake News Detection System In this project we propose a methodology to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine learning algorithms on an annotated (labeled) dataset that are manually classified and guaranteed. Then, feature selection methods are applied to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results. The product model will test the unseen data, the results will be plotted, and accordingly, the product will be a model that detects and classifies fake articles and can be used and integrated with any system for future use. Steps:

1.Download Dataset from Kaggle website.

2.Libraries required for the project.

- NumPy: NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. ↴
- pandas: Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. ↴
- Sklearn: Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms. ↴
- nltk: The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language.

3.Remove all non-words from news column.

4.Remove all stop-words using nltk. Stop Words: A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.

5. Then create features using TfidfVectorizer.

6. Then create X as input variable having feature and Y as output variable having label 0 for fake and 1 for real.

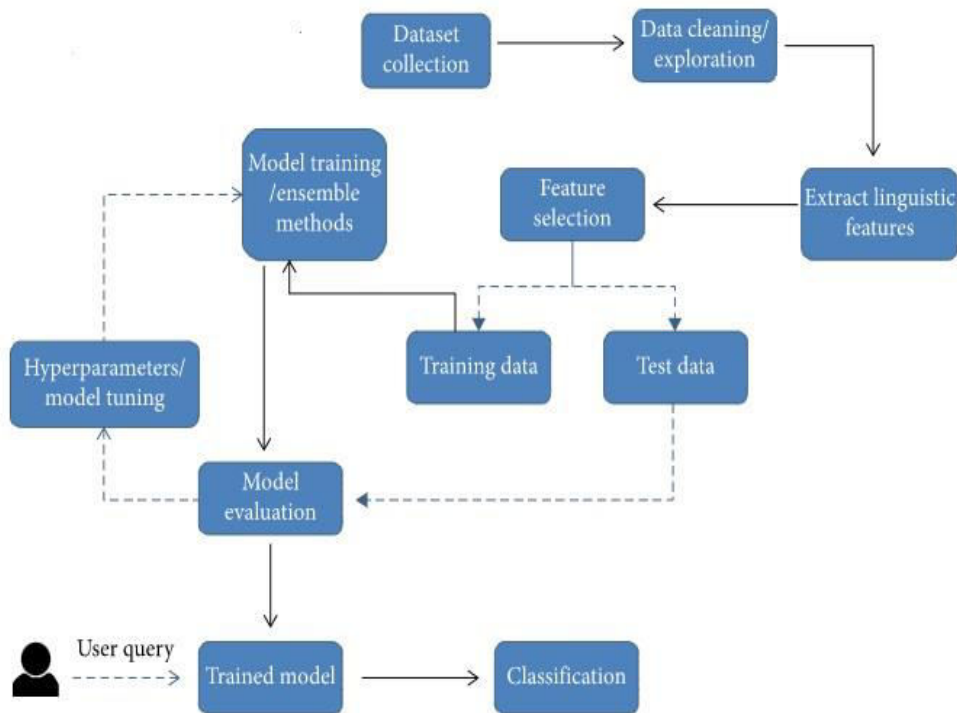
7. Then split data in training and testing.

8. Then pass training data to Passive Aggressive Classifier algorithm. Passive Aggressive Classifier belongs to the category of online learning algorithms in machine learning. It works by responding as passive for correct classifications and responding as aggressive for any miscalculation.

- Passive: If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.
- Aggressive: If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it.

9. Then we will check accuracy of model and deploy it on flask web framework.

10. Our web application will take a news as input and will classify it as Fake or Real according to model.



Proposed Methodology

IV. DATASET DESIGN

- We started by collecting a datasets 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 such as the ABC News, CNN, Times Of India, 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.

V. 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 (C_x) which x are believable and unbelievable. The probability of tweet data T in the class C_x can calculate as follow:

$$P(C_x|T) = P(T|C_x) \times P(C_x) / P(T)$$

- (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)
- (2) 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.

VI. TESTING

Our system take input from URL or an existing database & classify it to be real or fake news .To implement this various machine learning classifiers and a deep learning LSTM model have to be used.

VII. RESULT

- The news statement which is to be tested is entered in the textbox.
- The input data which is entered is processed through the proposed methodology and displays the output results whether it is true or false.
- The model must efficiently distinguish between Real news & fake news with Maximum Accuracy.

VIII. FUTURE SCOPE

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.

IX. CONCLUSION

It is significant to find the accuracy of news which is available on internet. In the paper, the components for recognizing Fake news are discussed. A mindfulness that not all, the fake news will propagate via web-based networking media.

REFERENCES

- [1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, —Fake News Detection on Social-Media: A Data Mining Perspective| arXiv:1708.01967v3 [cs.SI], 3 Sep 2017
- [2] Julio CS Reis, Andr´e Correia, Fabr´icioMurai, Adriano Veloso, Fabr´icioBenevenuto, and Erik Cambria. Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2):76–81, 2019.
- [3] M. Granik and V. Mesyura, "Fake news detection using naïve Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
- [4] Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017
- [5] Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys fake news.
- [6] Conroy, N., Rubin, V. and Chen, Y. (2015). —Automatic deception detection: Methods for finding fake news| at Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.
- [7] Markines, B., Cattuto, C., &Menczer, F. (2009, April). —Social spam detection|. In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web
- [8] Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP
- [9] Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, —Fake News Detection using Machine Learning and Natural Language Processing,| International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019
- [10] H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 380-383.
- [11] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, 2018, pp. 272-279.
- [12] C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," 2017 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, 2017, pp. 208-215.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



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