iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50344

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

A Review Tackling the Fake News Epidemic Using a Machine Learning Algorithm

Rahul Kumar Singh[1], Rajesh Bharati[2]
Department of Computer Engineering, Dr. D.Y. Patil Institute of Technology, Pimpri Pune 18

***Abstract:** Fake news has been a problem since the internet boom. Websites that keep us up to date with what's going on in the world are the perfect breeding ground for bad news and fake news. Fighting fake news is important because the world is knowledge-based. People do not make important decisions based on information; they also form their own ideas. Incorrect information can cause serious damage. It is not possible to identify all messages from a contact. This article attempts to speed up the fake news detection process by recommending a reliable fake news classification method. Machine learning contains different algorithms like naive Bayes, passive-aggressive classifiers and deep neural networks used eight different datasets from different sources. The text also includes the analysis and results of each model. With the right standards and the right tools, the task of detecting fake news will not be trivial.*
***Keywords:** Natural language processing; Classification; Logistic regression, Support vector machine, Fake news detection, TF-IDF.*

I. INTRODUCTION

A. Fake News

Fake news is a new phenomenon that has emerged with the growth of the internet and social media. In the past, traditional media monopolized the media, but the rise of the internet has liberated the media. While this had many positive effects, it also led to the growth of fake news, often designed to propagate propaganda or deceive people for financial gain. One of the main challenges in tackling fake news is that it can be difficult to distinguish between real and fake content. Fake news is often very persuasive, using wishful thinking and happy news to deceive people.

This can make it difficult to distinguish legitimate news from fake news. There are many different types of fake news, including hoax, clickbait, and propaganda. On the other hand, clickbait refers to sensational news designed to attract clicks, often with no mention of the truth of the story. Propaganda is another form of fake news designed to support an ideology or policy.

Detecting fake news is a complex process that requires many different methods. Finally, finding and combating fake news is a complex and ongoing process that requires cooperation between multiple countries. While there is no single solution to fake news, constant research and innovation is essential to solving this critical problem. especially. Fake news has become a major problem in today's society and many methods have been proposed to detect and combat it. In recent years, machine learning techniques have gained a lot of attention as they can learn patterns and properties of data.

II. LITERATURE SURVEY

Nair and Raut (2021) propose various methods to detect fake news using machine learning techniques. They discuss various extracti on methods such as word bag, ngrams, and word embedding and highlight the importance of feature selection and engineering. The review also covers various machine learning algorithms, including decision trees, support vector machines, and neural networks, an d evaluates their performance in terms of accuracy, precision, regression, and F1 score.

Thakur and Sharma (2020) provide a comprehensive review of different types of machine learning for fake news detection, includin g supervised, unsupervised, and deep learning. They compare the performance of these methods on different datasets and identify th e problems and limitations of the current approach. Here review also discusses importance of engineering selection and designs the need for translation and interpretation of machine learning models.

Abdulwahab and Al.Turaiki (2020) examine different methods of detecting fake news, including human fact checking, content analy sis and machine learning. They discuss the limitations of traditional methods and the advantages of machine learning, particularly in terms of scalability and automation. The analysis also highlights the need for information privacy and security and the importance o f social and cultural factors in detecting fake news.

Spinde et al. (2020) reports a multilingual method of fake news detection that combines natural language processing (NLP) with oth er models such as image and audio. The article argues that due to the complexity and diversity of fake news, it is difficult to detect f ake news through NLP alone, while multivariate techniques can provide better results and correction.
Horn et al. (2018) focused on searching for clickbait phrases using NLP techniques. This post shows how to monitor machine learni ng using various NLP features such as word frequency, sentiment analysis, and part of speech. The article also evaluates the perform ance of the model on different datasets and discusses the limitations and future aspects of using NLP to detect clickbait.

A. Limitations of Previous Studies

1) The data used for evaluation is not standardized, making it difficult to compare results from different studies.
2) The complexity and diversity of fake news makes it difficult to develop a model that can identify all types of fake news.
3) The quality of data used for training and evaluation can affect the performance of machine learning models.
4) Sharing the truth and interpreting machine learning models is still a challenge, especially in the context of fake news detection o f original interpretation.
5) Ethical and privacy issues related to the collection and use of user-generated content to teach educational standards should be addressed.

III.SYSTEM PIPELINE

The step by step process of completing the work of our process is as follows:

1) *Data Collection and Preparation:* Relevant data is collected from reliable sources, first of all to eliminate noise and irrelevant i nformation.
2) *Feature Engineering:* Extracting key features from previous data to detect fake news. This may include features such as word fr equency, sentence structure, purpose of analysis, and reliability.
3) *Model Selection:* Select the appropriate machine learning algorithm based on the characteristics of the data and the task at hand. Commonly used fake news detection algorithms include decision trees, logistic regression, support vector machines, and deep l earning models.
4) *Modeling:* Use domain data to train machine learning models to classify news as true or false. This includes feeding the model with pre-existing data and corresponding tags.
5) *Model Evaluation:* Evaluate the performance of a training model using classified data. This includes calculating metrics such as accuracy, precision, recall and F1 score.
6) *Distribution:* Use training models in real-world applications for real-time media distribution. This will involve creating an API that takes text messages as input and returns binary distributions.
7) *Continuous Improvement:* Monitor the effectiveness of the model used and continue to improve by re-implementing the model with new information and updating the architectural process.

IV. DESIGN

A. Data Collection and Preprocessing

Data collection and preprocessing are important steps in any machine learning process, including fake news detection. The quality a nd quantity of data used for training and evaluation can affect the performance of machine learning models.
The first step in data collection is to identify reliable and diverse sources of information. The data should contain a mix of real and f ake news in a few steps, including data cleaning, video extraction and data transfer.
Data cleaning includes removing irrelevant or unnecessary data, such as HTML tags or tags, and correcting typos and typos.
This step ensures that the data is consistent and ready for analysis.
Feature extraction is the process of transforming raw data into features that machine learning algorithms can use. In detecting fake news, features can include text such as word frequency, sentiment analysis and pattern analysis, as well as visual a nd auditory cues such as image metadata and audio recording.
Data transformation will transform the extracted features into a format that machine learning algorithms can use. This step includes t echniques such as normalization and scaling to ensure that the features are similar.
That allow machine learning models to be trained on representative examples. Social media platforms such as Twitter and Facebook are also important sources for detecting fake news as they provide access to user-generated content.

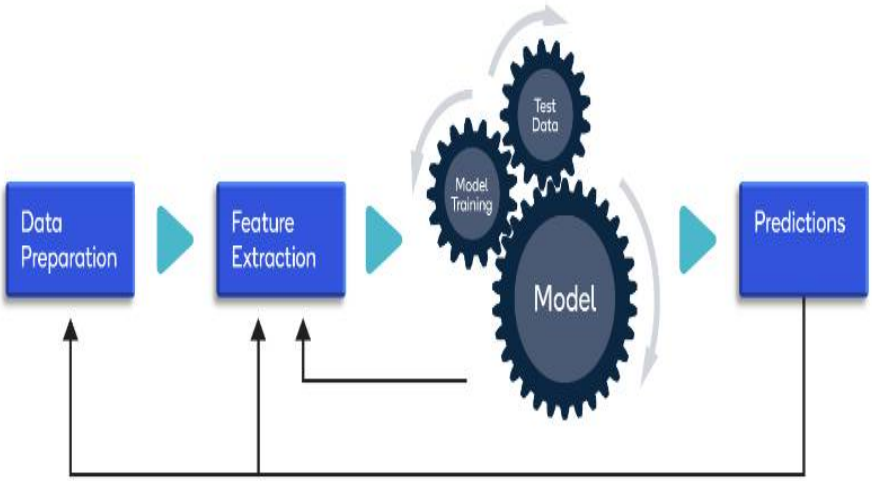


Fig. 1. Block Diagram

V. METHODOLOGY

Machine learning is one of the most powerful tools available today. In this article, we use pure machine learning to build our model. The task of choosing a classifier is due to the appropriate properties of the algorithm. Naive Bayes was chosen because of its simplic ity and robustness in class estimation, as it is popular in multi-
class estimation. In fact, one of the problems with other methods is that when new models are written, the model needs to be rerun to predict the results of the new data.
This is overcome by using a random operator that displays an incremental model, allowing changes only when needed, and discardin g updates when they do not change the equation.
We focus on learning models and problems based on deep learning. Deep neural networks are used to increase the efficiency of dete cting fake news. The following articles delve deeper into each algorithm.
The Naive Bayes classifier assumes that the features are statistically independent of each other.Specifies model properties as indepe ndent properties for a class. Because of their sense of freedom, they are highly skilled and can quickly learn to use advanced feature s with minimal training material. Given the data point $\vec{x}$ of n features, Naive Bayes predicts the class Ck of the data point, accordin g to Bayes' theorem,
can be decomposed to. Aggressive classifier due to its special properties. Passive- aggressive classifiers are simple and their performance has been shown to outperform
other methods. For example, the dataset where $\bar{x}$ is the data points and yi is the estimated output.

Given the weight vector w, the estimate can be easily obtained

$$\tilde{y}_t = \text{sign}(\bar{w}^T . \bar{x}_t),$$

Algorithms are generally used for this update method

$$\max(0, 1 - y_t(\bar{w}^T . x_t)) \qquad \frac{C}{||x||^2 + \frac{1}{2}} \qquad \bar{w}_{t+1} = \bar{w}_t + y_t \bar{x}_t$$

Therefore, this classifier informs the model incrementally. These are suggestions of algorithms that are not accurate compared to dee p learning architecture.

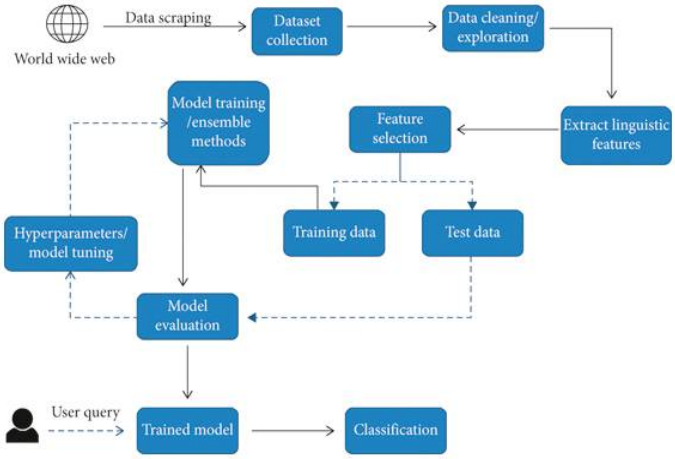


Fig2. Architecture Diagram

The Flow Chart [Figure 2] provides a brief overview of the entire model building process. The dataset is first cleaned for corrupt dat a and missing values
are removed. Some files have additional columns separated by relevance. Then the data is divided into training set, development set and test set. This model was refined through training and development and finally tested on a "test" set.

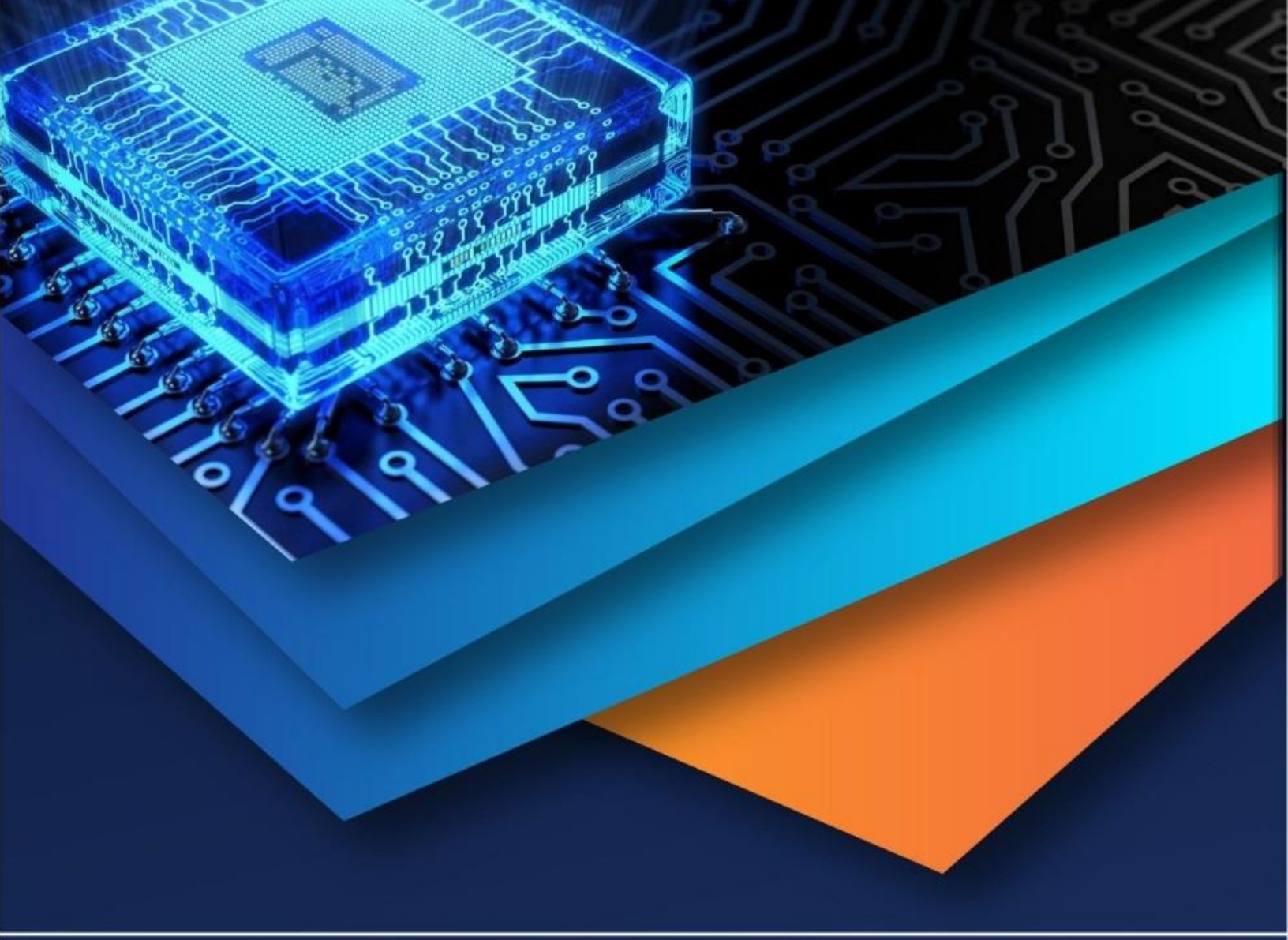
VI. CONCLUSION

our study demonstrates that machine learning algorithms have the potential to significantly improve the accuracy and speed of detecting fake news. This indicates that machine learning can be an effective tool for addressing the problem of fake news and improving the quality of information available to individuals and organizations. our study also highlights the need for further research in this area. Future studies should explore the use of larger and more diverse datasets and investigate the effectiveness of other types of features for detecting fake news, such as image-based features. Additionally, there is a need for continued efforts to educate the public about the dangers of fake news and the importance of verifying information before sharing it.

REFERENCES

[1] Khosla, N. K. Vaddi in P. K.Reddy, “Siv Machine Learning to Combat Fake News: A Review,” in Proceedings of the International Intelligent Sustainable Syste ms Conference, 2018.
[2] X. Li, X. Lu, thiab Q. Li, “A Deep Learning Approach to Detecting Fake News ", IEEE International Big Data Conference Proceedings, 2019.
[3] D. Dhanasekaran and S. Subramanian, "Detecting Fake News Using Machine Learning Techniques: A Systematic Review", Proceedings of IEEE International Conference on Intelligent Computing and Control Systems, 2019.
[4] A.Islam, R. Mishra and R. G. Ravichandran, "Detecting Fake News by Using Machine Learning Techniques: A Review," In Proceedings of the IEEE International Conference on Computational Intelligence and Computational Research, 2020.
[5] A. Alam, A.Gupta thiab S. K. Dwivedi, "Fake News Detection in Social Media: A Study of Methods and Trends", ACM Operations on Intelligent Systems and Technologies, vol. 12, no. December 5, 2021.
[6] A. Kumar and N. Rana, "Detecting fake news reviews using machine learning", Communications in Proceedings of the International Conference on Intelligence Computing and Emerging Trends , 2021.
[7] M.K. Paul, K. Paul, and N.N.R.Nair, "A comprehensive review of fake news detection using machine learning algorithms" in the Proceeding of the 2022, International Conference on Advances in Computing and Data Science.
[8] Atik Mahabub, "A powerful method for detecting fake news using survey sets and comparing it to other classifications".
[9] Map Reddy, Namratha Raj, Manali Gala, Annappa Basava, "Text mining-based detect fake news by the using the Ensemble method", International Journal of Automation and Computing.
[10] Sholok Gilda, "Information Analysis, Research Techniques for Analyzing Disruptive Information", IEEE 15th Student Research and Development Conference 2017.
[11] Hema, "Fake News Detection Using Machine Learning and Natural Language Processing", International Journal of Latest Technology and Engineering (IJRTE) ISSN: 2277- 3878 , nr 7, nr 6.

[12] Mykhailo Granik, Volodymyr Mesyura "Nrhiav Fake News detecting a Naive Bayesian Classifiers", 2017 IEEE 1st Ukrainian Electrical Engineering and Computer Engineering Conference.
[13] Akshay Jain, Amey Kasbe, "Fake News Detection", 2018 IEEE International Electrical, Electronics and Computer Science Student Conference.
[14] Samir Bajaj, "The Pope has a new baby! Use deep learning to catch fake news".
[15] A. Lakshmanarao, Y. Swathi, T.Srinivasa Ravi Kiran, "The Efficient Fake News Detection System Using Machine Learning", International Journal of Innovati ve Technology and Exploration Engineering, Vol. 8, no. August 10, 2019.
[16] Jru dataset available at kaggle: https://www.kaggle.com/jruvika/fakenews-detection#__sid=js0



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)