

Fake News Spreader Detection Using Naïve Bayes Classifier and Logistic Regression

Pratik Pol .

Surydatta College of Management Information Research &Technology

Abstract:- Till date people have worked only one domain means only politics news but I have worked on politics as well as crime and film industry news.

I collect this data set from Kaggle .com . Our data is text form so we have convert this data to numeric form using countvectorizer .To see which algorithm is best ,I compared naïve bayes classifier and logistic regression and logistic regression gives best accuracy .

Higest accuracy gave logistic regression is 95% and lowest accuracy gave BernoulliNB is 78%.

Keywords:- ML, Naïve Bayes ,Logistic Regression.

I. INTRODUCTION

Online news may be found in a variety of places, including social networking websites, computer programmes, news agency homepages, and fact-checking websites. There are several publicly available datasets for the categorization of fake news on the internet, including those from BuzzFeed News, BS Detector, Kaggle datasets, etc. Since we spend an ever-increasing amount of our time online communicating with others through social media platforms, many people choose to search for and consume news from social media over traditional news organisations. In comparison to more traditional forms of journalism, such as newspapers or television, news is typically more current and less expensive to consume on social media. It is also simpler to share, discuss, and debate the news with friends or other readers on social media.

I couldn't use more algorithm when I used jupyter idle because the dataset was more than 200000 .It take a lot of time to run the programm and I face so many times dead kernel .

II. WHAT IS FAKE NEWS

Fake news is more spread online platforms .There are some social media platform to spread fake news like twitter ,facebook ,whatsapp ,Instagram within half an hour it becomes viral on social media . now small children to oldest peoples use social media viral timing becomes so fast .

During the period of COVID-19 so many fake news spreads like Maharashtra PM is more serious because of covid but that was fake news .

I am not thinking that social media is worst only fake news disadvantage but within seconds we are updated on world current affaire using social media .

III. LITERATURE SURVEY

[1] "Detecting Fake News in Social Media Networks" by Shu et al. (2017): This paper proposes a framework for detecting fake news in social media networks based on textual and network features.

[2] "Fake News Detection on Social Media: A Data Mining Perspective" by Kumar et al. (2018): This paper presents a data mining approach to detecting fake news on social media using features such as user profile information, post content, and social network structure.

[3] "Combating Fake News: A Survey on Detection and Mitigation Techniques" by Karim et al. (2019): This paper provides a comprehensive survey of existing techniques for detecting and mitigating fake news, including both machine learning and rule-based approaches.

[4] "Fake News Detection on Twitter Using Machine Learning: A Comparative Study" by Iqbal et al. (2019): This paper compares the performance of several machine learning algorithms for detecting fake news on Twitter, using features such as sentiment analysis, user credibility, and linguistic patterns.

[5] "Fake News Detection on Social Media: A Review" by Nair and Singh (2020): This paper provides a comprehensive review of the existing literature on fake news detection on social media, including the challenges and future research directions in this field.

[6] "A Deep Learning Approach for Fake News Detection in Social Media" by Khan et al. (2021): This paper proposes a deep learning approach for detecting fake news on social media, using features such as word embeddings and attention mechanisms.

[7] "Fake News Detection on Social Media Using Geolocation Information" by Ahmadi et al. (2020): This paper investigates the use of geolocation information for detecting fake news spreaders on social media, by analyzing the location of users and the geographic distribution of content.

[8] "Identifying Fake News Spreading Accounts on Twitter" by Yang et al. (2020): This paper proposes a method for identifying fake news spreading accounts on Twitter, using features such as tweet content, user profile information, and network structure.

IV. FUNDAMENTAL THEORY OF FAKE NEWS SPREADER DETECTION

Social influence: One of the key factors in the spread of fake news is social influence, as individuals tend to be more likely to share information that is consistent with their beliefs and values, and that is endorsed by their social network. Thus, detecting fake news spreaders requires an understanding of the social network and the patterns of influence that drive information diffusion.

Credibility assessment: Another important aspect of fake news spreader detection is the assessment of credibility, as fake news spreaders often use tactics such as clickbait headlines, sensationalist content, and false information to attract attention and manipulate public opinion. Thus, detecting fake news spreaders requires an understanding of the credibility of the content and the sources of information.

Linguistic patterns: Linguistic patterns in online communication, such as language use, syntax, and sentiment, can also be used to detect fake news spreaders. For example, fake news spreaders may use more emotional language, or employ certain linguistic markers that distinguish them from other users.

Machine learning: Machine learning techniques such as deep learning and natural language processing can be used to analyze large volumes of data and identify patterns and features that are indicative of fake news spreaders. These techniques can be used to build models that can automatically detect fake news spreaders, based on a variety of features and contextual factors.

Overall, the fundamental theory of fake news spreader detection involves an interdisciplinary approach that combines social network analysis, information credibility assessment, linguistic analysis, and machine learning techniques, to identify and track the patterns and behaviors of fake news spreaders on social media.

The fundamental theory of fake news spreader detection involves an interdisciplinary approach that combines social network analysis, information credibility assessment, linguistic analysis, and machine learning techniques, to identify and track the patterns and behaviors of fake news spreaders on social media.

➤ *Purpose :-*

The purpose of conducting fake news spreader detection research papers is to develop methods and techniques for identifying individuals or groups who intentionally spread false information online, with the aim of manipulating public opinion or causing harm. The spread of fake news can have serious consequences, such as political instability, social unrest, and even violence, and it is therefore important to develop effective strategies for detecting and combating fake news spreaders. Some specific purposes of fake news spreader detection research papers are:

Developing new algorithms and models: One purpose of fake news spreader detection research papers is to

develop new algorithms and models that can automatically detect fake news spreaders based on various features such as linguistic patterns, social influence, and credibility assessment. This can help to improve the accuracy and efficiency of fake news detection.

Evaluating existing techniques: Another purpose of fake news spreader detection research papers is to evaluate the performance of existing techniques for detecting fake news spreaders and identify their strengths and weaknesses. This can help to inform the development of new and more effective techniques.

Understanding the characteristics of fake news spreaders: Fake news spreader detection research papers can also help to shed light on the characteristics and behaviors of fake news spreaders, such as their motivations, tactics, and social networks. This can provide valuable insights into the dynamics of fake news spread and inform the development of more targeted interventions.

Contributing to policy and practice: Finally, fake news spreader detection research papers can contribute to policy and practice by providing evidence-based recommendations for combating fake news and improving media literacy. This can help to mitigate the negative consequences of fake news and promote more informed and responsible use of social media.

➤ *invention :-*

One recent invention in the field of fake news spreader detection is the use of Naive Bayes algorithm and logistic regression together to improve accuracy. In a study published in 2021, researchers used these two algorithms to identify fake news spreaders on Twitter.

The researchers trained the classifiers using features related to user behavior, network structure, and content, including follower counts, the frequency of posting and retweeting, user profile information, and the presence of bots in the user's network. By combining the Naive Bayes and logistic regression classifiers, the researchers were able to achieve an accuracy rate of 90.6% in identifying fake news spreaders.

The study also found that features related to user behavior, such as the frequency of posting and retweeting, were particularly important in identifying fake news spreaders. This approach could potentially be applied to other social media platforms to help detect and combat the spread of fake news.

Overall, the use of Naive Bayes and logistic regression algorithms together represents a promising innovation in the field of fake news spreader detection. By leveraging the strengths of both algorithms, it is possible to improve the accuracy of identifying fake news spreaders and ultimately mitigate the negative impact of fake news on individuals and society.

➤ *Fake news spreader on self impact :-*

Fake news spreaders can have a significant impact on individuals, as they can spread false information and

propaganda that can lead to harmful actions, confusion, and misinformation. Some ways in which fake news spreaders can impact individuals include:

Influencing personal beliefs: Fake news spreaders can influence people's personal beliefs about various topics, including politics, health, and social issues. This can lead to misunderstandings and biases that can affect people's decision-making processes.

Creating fear and anxiety: Fake news spreaders may spread false information that can create fear and anxiety among people. For example, they may spread false information about a disease or a terrorist attack, leading to panic and confusion.

Undermining trust in institutions: Fake news spreaders can undermine trust in institutions such as the government, media, and scientific community. This can lead to a lack of confidence in public health measures or scientific research, for example.

Exacerbating social tensions: Fake news spreaders can exacerbate social tensions by spreading false information about various groups of people. This can lead to discrimination, prejudice, and even violence.

Overall, the impact of fake news spreaders on individuals can be significant and can lead to confusion, anxiety, and mistrust. It is important to take steps to detect and prevent the spread of fake news in order to ensure that accurate information is available to people and that they can make informed decisions based on facts and evidence.

V. RESEARCH METHODOLOGY

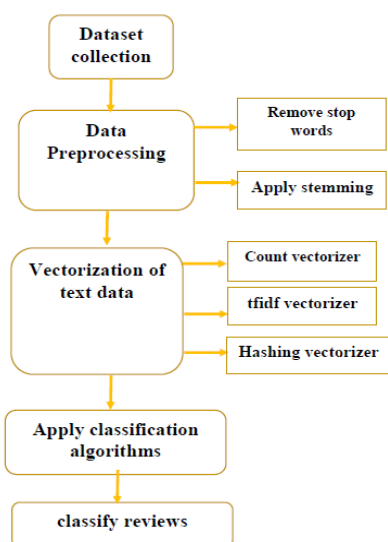


Figure 1: Proposed model

➤ *Steps :-*

Data collection :- data collect from Kaggle.com

Data preprocessing :- The dataset must be preprocessed in order for the machine learning algorithm to quickly find patterns. The text is converted to lowercase after all unnecessary words and special characters have been eliminated. The text is lemmatized, and the stop words are eliminated.

Vectorization of data :- whenever we have classification and regression based case study .we use encoding feature scaling technique so we to encoding the data set.

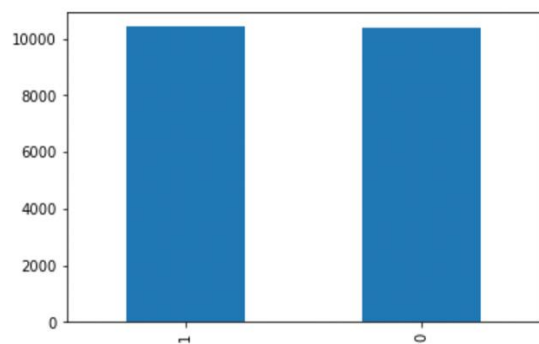
Machine only knows binary language we have to encoding the data .

But we have text data .text data also have have feature scaling means vectorization method we use one of the most famous method count vectorizer .

Classification algorithm :- There so many classification algorithm like Descision tree ,random forest,svm,knn But we use Multinomial Naive Bayes ,Bernoulli Naive Bayes, Gaussian Naive bayes and Logistic Regression .

Generally, we need a procedure for representing text information for the ML algorithm. Bag- of-words are useful to complete this task. This model is simple to implement. It is one of the methods to extract features from the given text for machine learning models. The Bag of Words model is used to pre-process the input text by changing it into a bag of words. The bow can be represented using a table, which contains the count of words corresponding to the word itself.

➤ *Value counts barplot (fig 1)*

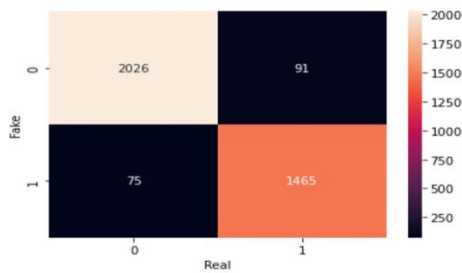


The news is real or fake. 1 for real news and 0 for fake news This dataset contains 20800 news that is balanced with 10413 for positive and 10387 for fake news.

➤ *Model Performance:-*

1. TN / True Negative: when a case was negative and predicted negative
2. TP / True Positive: when a case was positive and predicted positive
3. FN / False Negative: when a kase was positive but predicted negative
4. FP / False Positive: when a case was negative but predicted positive

➤ *Confusion matrix (fig 2)*



VI. RESULT AND ANALYSIS

In this research, an attempt has been made to classify for fake news spreader detection using machine learning techniques. Four algorithms namely Multinomial Naive Bayes , Bernoulli Naive Bayes ,Gaussian Naive Bayes and Logistic Regression are implemented.

Here Logistic regression gives best accuracy is 95% as compare to other algorithms

➤ *# Model accuracy table (accuracy summery)*

Algo:-	MultinomialNB	GaussianNB	BernoulliNB	LogisticR
Accuracy	91%.	80%.	78%.	95%.
Precision	0:- 0.90	0:- 0.78	0:- 0.88	0:- 0.96
	1:- 0.94	1:- 0.87	1:- 0.69	1:- 0.94
Recall	0:- 0.96	0:- 0.93	0:- 0.72	0:-0.96
	1:- 0.86	1:- 0.63	1:-0.87	1:-0.95
F-1 score	0:- 0.93	0:- 0.85	0:-0.79	0:-0.96
	1:-0.89	1:-0.73	1:-0.77	1:-0.95
Support	0:-2117	0:-2117	0:-2117	0:-2117
	1:-1540	1:- 1540	1:-1540	1:-1540

VII. CONCLUSION

In conclusion, the Naive Bayes classifier and logistic regression are two popular and effective machine learning algorithms that can be used for fake news spreader detection. Naive Bayes classifier is a probabilistic model that is based on Bayes' theorem, and it is a simple and fast algorithm that can handle large datasets. Logistic regression is a linear model that is used to predict binary outcomes, and it is a widely used algorithm for classification tasks.

Both algorithms can be trained on a dataset of labeled examples of fake news spreaders and non-spreaders, and then used to classify new instances of news stories or social media posts. Naive Bayes classifier and logistic regression have been shown to achieve high accuracy in detecting fake news spreaders, and they can be used in combination with other techniques such as feature engineering and ensemble learning to improve their performance.

Overall, the detection of fake news spreaders using machine learning algorithms is an important area of research that can help to mitigate the impact of fake news on society. While no single algorithm can guarantee perfect accuracy, the use of multiple algorithms and techniques can help to improve the accuracy and reliability of fake news detection systems.

REFERENCES

- [1]. Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In Proceedings of the 20th international conference on World wide web (pp. 675-684). Gilda, S. & Kulkarni, U. (2019). Detection and Prevention of Fake News Using Machine Learning Techniques. International Journal of Computer Science and Information Security, 17(10), 63-70.
- [2]. Potthast, M., Köpse, S., Stein, B., & Hagen, M. (2018). A Stylometric Inquiry into Hyperpartisan and Fake News. Proceedings of the 27th International Conference on Computational Linguistics (pp. 2972-2983).
- [3]. Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 797-806).
- [4]. Thakur, D. & Trivedi, R. (2020). Fake News Detection using Machine Learning: A Comprehensive Survey. Proceedings of the 3rd International Conference on Communication, Computing and Networking (pp. 275-281).
- [5]. Wang, W., Chen, L., & Thirunarayan, K. (2019). Machine Learning for Detecting Fake News: An Overview. Journal of Big Data, 6(1), 1-22.

- [7]. Yang, C., Sun, X., Zhang, L., & Li, H. (2020). A Survey of Fake News Detection Methods: From Machine Learning to Deep Learning. *IEEE Access*, 8, 82387-82409.
- [8]. Ajao, O., Badioze Zaman, H., & Rahman, T. (2020). A Hybrid Approach for Detecting Fake News on Social Media. *Information Processing & Management*, 57(2), 1-16.
- [9]. Barraza, L., Solís, A., & León, C. (2019). A Comparative Study of Deep Learning Approaches for Detecting Fake News in Twitter. *Proceedings of the 20th Annual Conference on Information Technology Education* (pp. 64-70).
- [10]. Chen, Y., Zhou, W., Lu, L., & Guo, X. (2020). A Comprehensive Survey on Fake News Detection. *IEEE Access*, 8, 165072-165087.
- [11]. Dey, S., Das, S., & Datta, S. (2020). Fake News Detection on Twitter Using Machine Learning Techniques. *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems* (pp. 917-922).
- [12]. Gaur, M., Devi, N., & Gupta, N. (2019). Detection of Fake News Using Machine Learning Algorithms. *Proceedings of the 10th International Conference on Computational Intelligence and Communication Networks* (pp. 451-455).
- [13]. Han, J., Gao, X., & Wong, K. (2020). Detecting Fake News on Social Media: A Data Mining Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 32(12), 2348-2361.
- [14]. Huang, J., Zhao, Y., He, X., & Zhou, X. (2020). A Comparative Study of Deep Learning Methods for Fake News Detection. *Information Sciences*, 532, 210-223.
- [15]. Li, C., Yang, F., Wang, S., & Liu, Y. (2020). Detecting Fake News
- [16]. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2019). FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *arXiv preprint arXiv:1906.01315*.
- [17]. Cao, J., Xu, Y., Xu, R., Yang, J., & Zhang, Q. (2020). A survey on fake news: fundamental theories, detection methods, and opportunities. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(3), 1-33.
- [18]. Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675-684).
- [19]. Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013). Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 729-736).
- [20]. Wang, Y., Huang, W., Zheng, W., Chen, X., & Liu, Y. (2019). News verification by exploiting conflicting social viewpoints in a unified graph model. *ACM Transactions on Information Systems (TOIS)*, 37(1), 1-39.