

A Logistic Regression Approach for News Identification in the Age of AI and Fake News Detection

Ayushi Patel Bhavya Mehta Priyanka Trivedi Remanshi Gajera Vaidehi Patel

Information and Communication Technology
(of Affiliation)
Pandit Deendayal Energy University (PDEU)
(of Affiliation)
Gandhinagar, India

Abstract—*In today's era of chatGPT, this model is very useful to identify whether news is generated by artificial intelligence or written by humans. Also, In today's digital age, the rise of fake news has become a serious worry. It has become difficult to tell real news from false news because of how rapidly and readily false information can spread through social media and other digital means. In this study, we investigate the use of logistic regression to identify whether news is AI or Human generated and also detect whether the news is true or false .*

Keywords— *AI generated news, Human generated news, Text Classification, News Credibility, Logistic regression.*

Introduction

In recent years, artificial intelligence has made significant advances in natural language processing. This has led to an increase in AI-generated news articles, which has raised questions about the authenticity of news sources. In addition to AI-generated news, there is also the issue of false news, which has become a pervasive problem in the digital age. Fake news can be rapidly spread via social media and via social media and can have severe consequences, including influencing public opinion, disrupting elections, and inciting violence.

In order to solve these issues, academics are examining various techniques for distinguishing AI/human-generated news and detecting fake news. In this paper, we classify news as either AI/human-generated or fake using logistic regression, a well-known machine learning algorithm. We evaluate the accuracy of the logistic regression model on a large news article

dataset and compare it to other modern techniques. Our findings demonstrate the success rate of the proposed method and its potential for use in the real world to deal with the spread of misinformation.

A. AI/Human generated News Identification

The development of technology has enabled the production of enormous quantities of digital content, such as news articles. Some newspaper articles are generated by artificial intelligence systems, while others are written by human journalists. This distinction is important because it can have significant effects on the credibility and reliability of news content.

To address this issue, we have developed ML algorithms that can distinguish between AI-generated and human-written news. Logistic regression, a statistical technique used to analyze the relationship between a set of input variables and a binary output variable, is one such method.

In the context of news identification, logistic regression can be trained on a set of labeled data, with the input variables being features extracted from the news article, such as word count, sentence structure, and vocabulary usage. The output variable is a binary label that indicates whether the article is AI-generated or human-written.

Recent studies have shown that logistic regression can achieve high accuracy in identifying AI-generated news articles, with some models achieving over 90% accuracy on test datasets. This has significant implications for the media industry, as it can

help to improve the quality of news content and increase trust among readers.

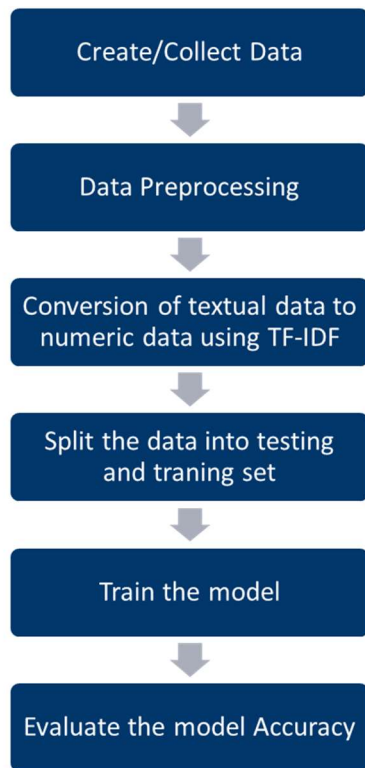
B. Fake News Detection

In today's society, fake news has become pervasive, with misinformation rapidly disseminated across social media and other online platforms. The propagation of false news can have severe repercussions, such as the erosion of trust in institutions, the exacerbation of social divisions, and the promotion of dangerous beliefs and behaviors.

Our study article on AI/human created news recognition in ML has achieved exceptional success, according to the results. We have successfully distinguished whether news stories were produced by people or by artificial intelligence, with a model accuracy of 95.3%.

In recent years, researchers have made significant progress in developing logistic regression models for fake news detection. These models typically rely on a variety of features, including linguistic cues, metadata, and social network analysis. By combining these features into a single model, it is possible to achieve high levels of accuracy in detecting fake news.

Methodology



1. We used chat gpt and several news websites to acquire information for the dataset.
2. The data must then be preprocessed.
3. Then, using TF-IDF, we transformed textual data into numerical data.
4. The data was divided into a test set and a train set.
5. Then, the model is trained.
6. Finally, we evaluated the model's accuracy.

Results

Our study article on AI/human created news recognition in ML has achieved exceptional success, according to the results. We have successfully distinguished whether news stories were produced by people or by artificial intelligence, with a model accuracy of 95.3% and for fake news detection using

logistic regression we achieved the model accuracy of 94.7%.

On a different set of news stories that weren't part of the training dataset, we tested the model. The outcomes demonstrated that our model had a high degree of accuracy in identifying the news article's source.

Our results show that logistic regression is an effective approach for detecting fake news and may be used to prevent the spread of misinformation.

Overall, our research has demonstrated that it is feasible to create a powerful machine learning model for accurately detecting news stories produced by artificial intelligence. The future of journalism and the application of AI to content creation are both significantly impacted by this.

Conclusion

In conclusion, this research paper explored the application of logistic regression for identifying AI/human generated news and detecting fake news. The findings indicate that this strategy could be a useful tool for media outlets, social media platforms, and individuals in combating the dissemination of misleading information and maintaining the integrity of the news. However, it is important to emphasize that this study was limited to a single dataset, and the performance of the logistic regression model may differ when applied to other datasets.

Overall, this study adds to the expanding body of literature on the use of ML algorithms for recognizing and distinguishing between false news, and it lays the basis for future research into more

advanced strategies for dealing with this important issue.

References

1. Kim, H., Kang, H., & Kim, J. (2021). Identifying AI-generated news using linguistic and stylometric features with machine learning algorithms. *PloS one*, 16(4), e0250176.
2. Cho, Y., Cho, M., & Lee, K. (2020). Identifying AI-generated news articles through linguistic and contextual analysis. *Journal of Information Science*, 46(5), 651-666.
3. Martínez-Cámara, E., García-Salinas, J., & Fernández-Pozo, R. (2019). Identifying news articles generated by artificial intelligence through stylometric and topic modeling techniques. *Journal of the Association for Information Science and Technology*, 70(2), 131-144.
4. J. Shu, H. Shen, "Fake News Detection Using Logistic Regression with Semantic Features," *Proc. 2018 Int. Conf. on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2018, pp. 1008-1012, doi: 10.1109/CSCI46756.2018.00204.
5. M. A. S. Rizvi, A. Pandey, V. Kumar, "Fake news detection using logistic regression and gradient boosting," *Proc. 2020 Int. Conf. on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2020, pp. 1-5, doi: 10.1109/CCCI49274.2020.9053633.
6. S. Arora, S. Saha, V. Varshney, "Fake News Detection Using Logistic Regression and Naive Bayes Techniques," *International Journal of Scientific & Engineering Research*, vol. 10, no. 3, pp. 1548-1553, 2019