Fake News Detection Using Machine Learning

Vasavi Kullanakoppal

Bellevue University

DSC 680: Applied Data Science

Prof. Matthew Metzger

June 20, 2025

Fake News Detection Using Machine Learning

Business Problem

Misinformation and disinformation have proliferated widely in internet news, eroding public trust and influencing opinions. The business problem is to come up with a trusted and scalable machine learning system to autonomously detect fake news articles to support content moderation, journalism integrity, and public literacy.

Background/History

Increased social media and web platform presence has enabled instantaneous news broadcasting without official editorial oversight. Publicized incidents of fabricated news influencing political outcomes, health decisions, and social movements have spurred scientific and technological response. Traditional fact-checking processes are not suitable for scaling up; hence, AI-based systems are a feasible alternative.

Data Explanation (Data Prep/Data Dictionary/etc)

The data used in this study consists of thousands of labeled news, categorized as either Fake or Real. Preprocessing has been used on text data using tokenization, stop word removal, punctuation, and special character removal. Other characteristics such as word count from an article were also derived to understand structural variations. The bulk columns in the data set were text, consisting of the whole article body; label, a binary attribute with 0 denoting Fake and 1 denoting Real; and publication date, which was applied to analyze trends in articles through time. The cleaned and structured data set formed the basis for testing and model training.

Methods

Supervised learning was applied to categorize news articles in this project. The information was split into an 80% training set and a 20% test set. Text information was vectorized using Term Frequency-Inverse Document Frequency (TF-IDF), a popular way of representing text information as numerical values. Several classification models were trained and tested, starting with Logistic Regression as a baseline. Metrics for analysis were accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC) that all together provided a holistic view of model performance.

Analysis

The exploratory data analysis (EDA) indicated that the dataset was well balanced between Fake and Real news articles. Fake news stories were also typically shorter than Real news, and most stories were published in 2016, suggesting a peak that can be correlated with social or political activities. The previous machine learning model had 98% accuracy, with precision, recall, and F1-scores for both classes all at 0.98. The ROC curve had an AUC of 1.00, indicating excellent separability between the Real and Fake classes. These results highlight the stability and deployment ability of the model in real-world applications.

Conclusion

The model for detecting fake news exhibited excellent performance, both in terms of predictions and impartial treatment of the two classes. The test showed not just that the model was capable of classifying articles precisely with high accuracy but also that it possessed a very low false positive and false negative rate. The ability of the model to distinguish between false news articles and actual news articles is nearly flawless with an AUC of 1.00. These results show that the model is practical and trustworthy for use in computerized content verification systems.

Assumptions

This study presumes that all the labels in the dataset are valid and indicative of the content of the article. It also presumes that the text field represents the entire article, and that article's length is indicative of complexity or genuineness. These presumptions are necessary for the model to learn from patterns in data.

Limitations

Although it has done a good job, the model is not flawless. It likely won't do quite as well on casual writing such as users' comments or social media posts. The model is only trained on English-language news articles from 2015 to 2017, so its usefulness on future or multilingual content is questionable. Op-eds and satire, commonly written in a comparable style to misleading news, can also prove to be challenging.

Challenges

Some of the most prominent challenges include avoiding prejudice when model training, specifically in processing politically or culturally sensitive material. Satire, parody, or opinion-driven news articles are still challenging machine learning algorithms. The dynamic nature of disinformation is another major challenge because it requires frequent updates in models to remain current.

Future Uses/Additional Applications

There are many potential applications for this model, which can be used in content management systems (CMS) or browser extensions as instant article classification. It can also be adapted for use with multiple languages or used in schools to help users critically evaluate news. In the future, integrating it with fact-checking APIs and knowledge graphs could make it even more useful.

Recommendations

To maintain performance, the model needs to be repeatedly retrained with new data. As a precaution, it is also recommended to enrich the dataset with additional features, such as authorship or credibility. Inclusion of user feedback into the classification can ensure ongoing performance improvement. Finally, collaboration among stakeholders is important to ensure ethical deployment and transparent reporting of model predictions.

Implementation Plan

The deployment has a number of phases. Initially, the model needs to be finalized with the current version of the dataset. Next, it needs to be scaled on a real-time prediction-capable cloud. A monitoring system has to be in place to keep track of model performance in production. Lastly, a scheduled retraining process needs to be put in place to manage data drift and maintain high accuracy.

Ethical Assessment

Application of fake news detection systems must consider ethics of transparency, accountability, and abuse. The system must provide confidence scores and assist with human oversight, particularly in close cases. The system must be trained using broad data sets to avoid inbuilt bias and not used in censoring legitimate journalism. Compliance with data privacy acts like GDPR must be honored, particularly if it is handling user-input content.

10 Audience Questions

1. How well does the model's accuracy hold up in real-world deployment?

2. Can it detect satirical or parody news?
3. How often does the model need to be retrained?
4. Is the model biased towards certain sources?
5. Which of the features contribute the most to predictions?
6. Can the system be gamed or bypassed?
7. How well does it handle multilingual text?
8. Can explanations be viewed for all predictions?
9. What are the ethical safeguards taken?
10. How scalable is this solution for enterprise deployment?

Reference:

Bisaillon, C. (n.d.). Fake and Real News dataset. *Kaggle*.
https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset