

Sentiment Analysis of Fake/Real News

A DS 4002 Case Study by Kyle Tran



How do you know if you can trust a news article you read? Can you simply accept the information at face value, or do you have a responsibility to verify the reliability of your sources? According to Pew Research [1], about 32% of Americans report frequently encountering fake political news, contributing to widespread misinformation and growing distrust across the political spectrum. With the rise of artificial intelligence (AI), it has become even easier for false information to spread. Social media platforms, in particular, continue to be a major vehicle for the dissemination of fake news. Given these challenges, this case study will examine the role sentiment plays in the spread of fake news.

An important part of identifying fake news involves understanding whether sentiment patterns can help distinguish fake articles from real ones. If a relationship between sentiment and fake news exists, it could offer valuable insights for helping the public better recognize misinformation.

In this project, you will take on the role of a data scientist tasked with investigating whether there is a significant difference in sentiment between fake and real news articles, and whether this difference can be used to distinguish between them. Specifically, you will answer the research question: Does the sentiment score of news articles tend to be more negative for fake news compared to real news, and does this negativity increase as the 2016 presidential election approaches?

To explore this question, you will apply sentiment analysis techniques — using at least the VADER Python package — to assess the sentiment of various political news articles published around the 2016 U.S. Presidential election. (If you are unfamiliar with VADER, you can find a brief overview here [2].) Finally, you will use a statistical method of your choice (such as a t-test, Mann-Whitney U test, etc.) to determine whether the sentiment differences between real and fake news are statistically significant. More details can be found on the rubric.

[1] M. B. Holcomb Amy Mitchell and Jesse, "Many Americans Believe Fake News Is Sowing Confusion," Pew Research Center. Accessed: Apr. 27, 2025. [Online]. Available: <https://www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>

[2] L. Geetha, "Vader: A Comprehensive Guide to Sentiment Analysis in Python," Medium. Accessed: Apr. 27, 2025. [Online]. Available: <https://medium.com/@rslavanyageetha/vader-a-comprehensive-guide-to-sentiment-analysis-in-python-c4f1868b0d2e>