Predicting Fake and Real News

Student : Jasper Zhou **Instructor:** Dr. Adam Andersen
DIGHUM 100, SUMMER 2021

Research Questions:

A) How has fake news changed?
B) Can we predict whether something is fake news or not using NLP?

Introduction:

Fake news is not a recent or unique issue, although it has been exacerbated by the ease and speed in which information could be spread. Since the 1890s, fake news has plagued the profession of journalism. Dubbed "yellow journalism", newspaper publishers would run sensationlistic news articles. However, fake news was eventually replaced by objective journalism, until recently due to the rise of the internet. (Source:UCSB).

In terms of impacts fake news has on society, University of Michigan lists the following as possibilities:

- Anti-intellectualism
- Antiscience
- Widespread Mistrust

EDA

-Explore data available
-Describe data here
-Ideas:
    -Average length (word count) in fake news vs. real news (Grouped bar chart) vs. time
    -Word Cloud of "Real" and "Fake News"
    -Track certain key words in word cloud
-Also, explanations below each image

Tracking Changes in "Fake" and "Real" News:

-Using Sentiment Analysis Tools (TextBlob)
-Track polarity and subjectivity of:
    -Fake News
    -Real News
    -Difference between fake and real news in both subjectivity and polarity
    -All vs. time
-Insert explanation of Sentiment Analysis results

Predicting Fake News vs. Real News:

-Explain model fitting process:
    -Logistic Regression
    -SVM
    -etc.

-Analysis of performance of models

-ROC Chart/ Binary Confusion Matrix

Conclusion:

-Summarize findings of EDA
-Summarize changes between fake and real news
-Summarize findings from predictive task
-Analyze suitability of tools used for task

Works Cited

-Insert citations: articles, documentation, etc.