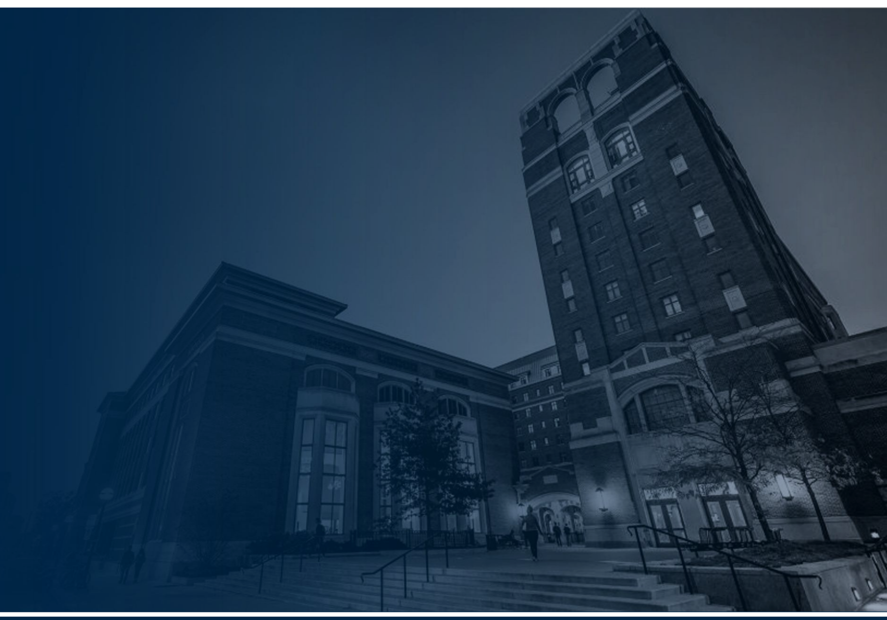
When fact is false and false is funny:

Automated detection of fake news and satire

Christine Gregg, Liwen Huang, & Ali Tobah | University of Michigan, Ann Arbor



BACKGROUND

*The Universe of News**

* Universe and content not to scale

False information comes in different forms, and for many reasons. In the study of social media, Big Data, and the Internet, three names for false information are common: misinformation, disinformation and fake news. In this study, we use 'fake news,' because we focus on false information in news articles without regard for intent.

For our purposes, we are considering a simplified concept of the "universe of news" where there are only three categories: real news, fake news, and satirical news. Our focus is on identifying the features that have the greatest impact on differentiating these types of news with the hope of making the distinction between them more clear for everyday news readers.

MOTIVATION

Can we translate patterns in news articles identified by machine learning models into useful "red flags" for news readers that struggle to differentiate between real, fake, and satirical news?

METHODS

Data Sources

We worked with three binary comparison data sources containing labeled real, fake, or satirical news text:

- Real-Fake: Pérez-Rosas (2017)
- Real-Satire: Yang et al. (2017)
- Satire-Fake: Golbeck et al. (2018)

Feature Engineering

To maintain comparability across models, our goal was to use similar features at the start and add specific features identified in the literature as-needed according to model performance. All datasets contained the full body of article text vectorized into terms with the Scikit-learn TF-IDF vectorizer. Additional features used in select models include:

- Article title (TF-IDF vectorized)
- Punctuation (as percentage of characters)
- Linguistic features (identified with Empath)

Classes were relatively balanced with the exception of the Real-Satire dataset, where the majority class was downsampled. Imbalance was also addressed at the article and sentence length level in this dataset.

Model Selection

We tried a number of models including Logistic Regression, Linear SVC, Multinomial Naive Bayes, and Random Forest.

RESULTS

Model Performance

In each binary comparison case, Logistic Regression provided the best results, as defined by maximizing the accuracy and F1 Score (**Table 1**). The most successful combination of features was used in **Table 1**.

Table 1: Logistic Regression Results

Model	Accuracy	F1 Score
Real vs Fake	0.76	0.73
Real vs Satire	0.94	0.96
Satire vs Fake	0.78	0.70

Feature Importance

Logistic Regression feature weights for each model are provided in **Figures 2**, **3**, and **4**. Some make intuitive sense, such as official-sounding, yet unspecific, features like "insider", "source" or "many" for fake news.

Figure 2: Real | Fake Feature Importance

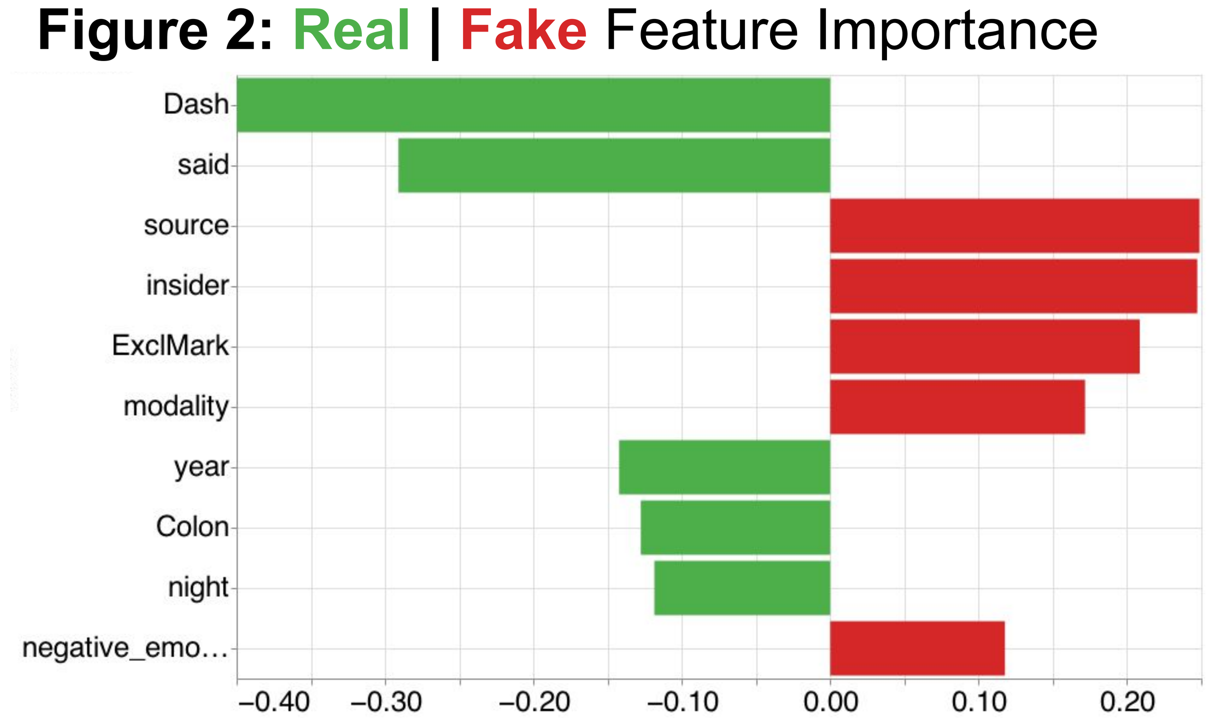


Figure 3: Satire | Real Feature Importance

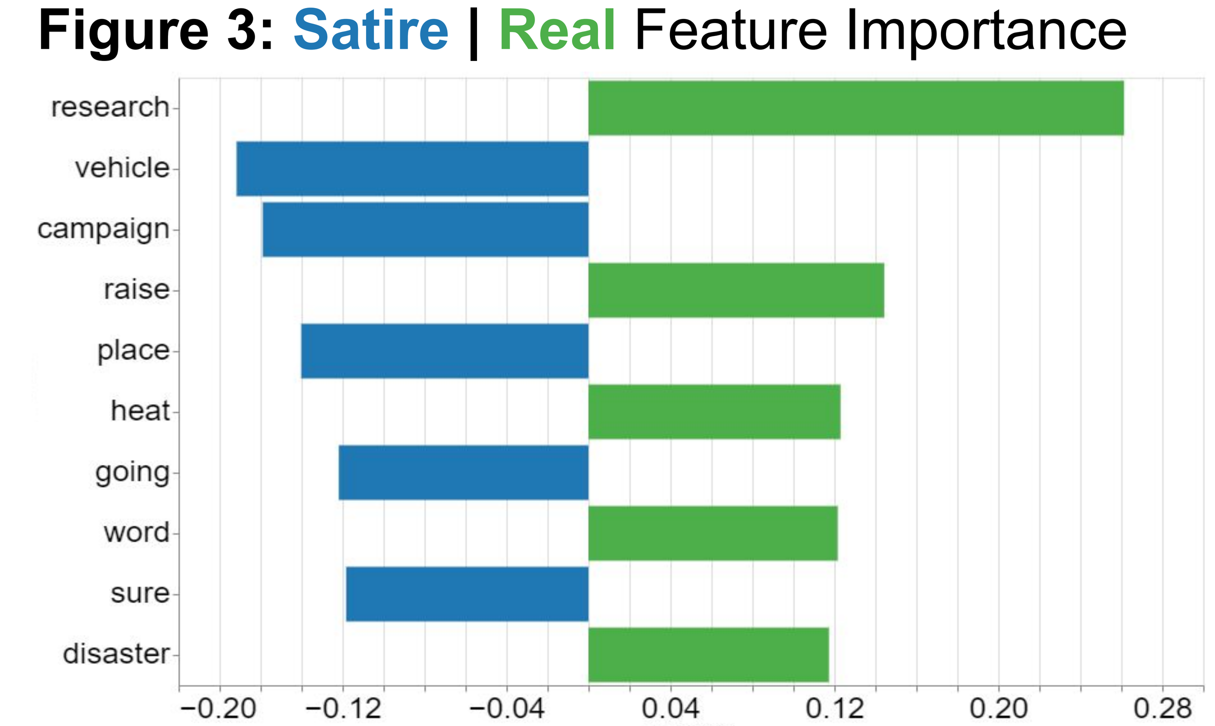
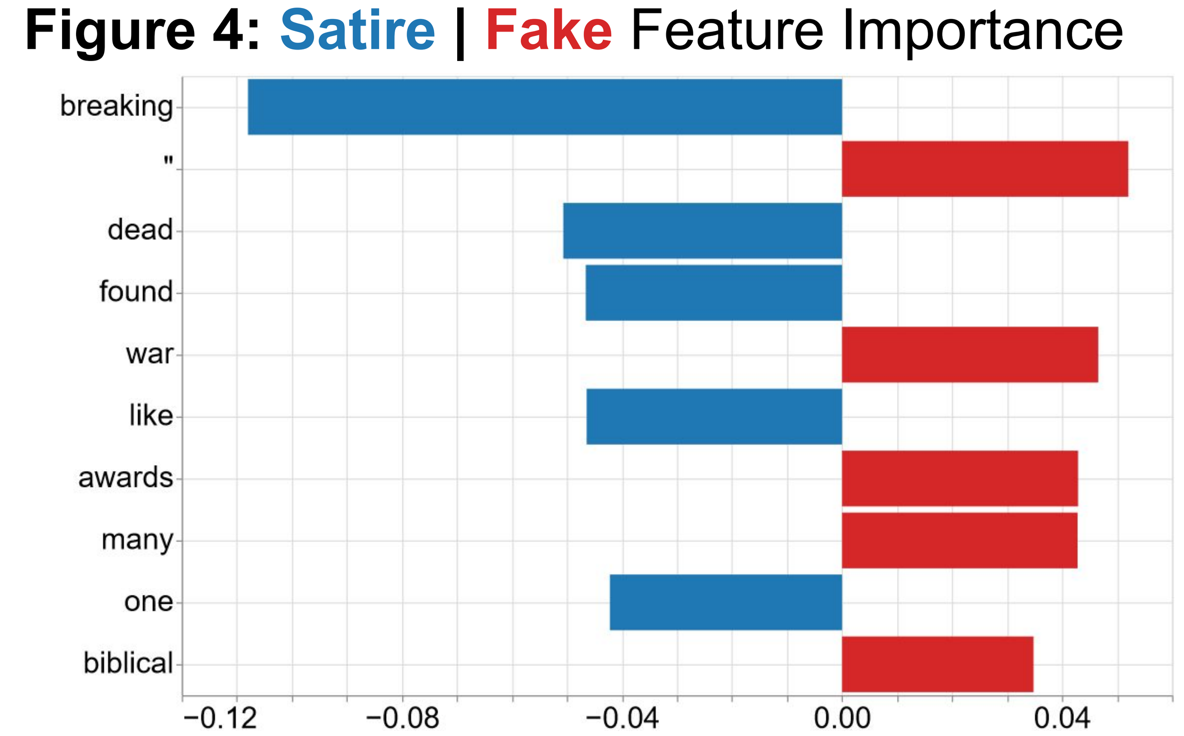


Figure 4: Satire | Fake Feature Importance



DISCUSSION

Logistic Regression distinguished between real and satirical news successfully, but struggled to do so for real and fake or satirical and fake news. A simple reason could be dataset related: the real-satire dataset was larger. However, the classifier does give intuitive and explainable results for the two lower-scored comparisons.

Further Considerations

Among other limitations and ethical pitfalls described in our full report, news classifiers should be routinely updated with samples of the latest news that reflects a variety of sources, topics, and cultural contexts.