DETECTING FAKE NEWS

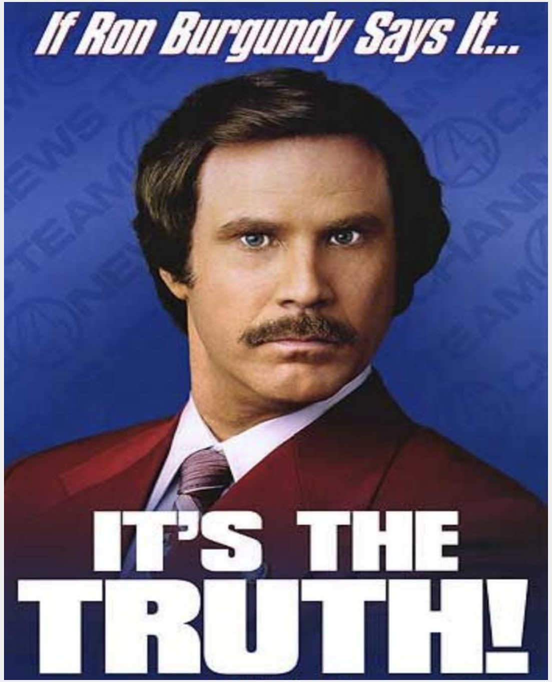
Benjamin Karlsberg

THE PROBLEM

Can we trust what we read on the internet?

Many actually benefit from spreading false information

How can we try to mitigate this issue?



MY APPROACH

ACQUIRE MANY FACT-CHECKED ARTICLES

USE A TF-IDF VECTORIZER TO CREATE A "BAG OF WORDS"

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

TEST SEVERAL MODELS TO DECIDE BEST FIT FOR DETECTION

DATA INFORMATION

Article data sourced from Politifact.com, The New York Times, and The Onion Magazine (Satire)

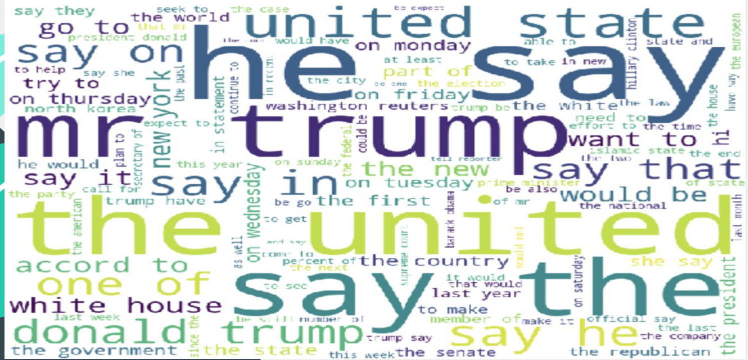
- Dataset:
    - > 70,000 articles
    - Class Balance: 47% Fake, 53% Real
    - 10% NYT, 10% Onion, 80% Politifact Data (Only political topics)

- TF-IDF:
    - Lemmatized words
    - Includes Bigrams and Trigrams of words
    - 20,000 Feature Limit

Fake vs Real Articles

Common words from "Fake" articles

Common words from "Real" articles



"Fake" Examples:

Politifact: 'Donald Trump just couldn't wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out to his enemies…'

Onion: 'Man Wasting His Life Playing Video Games When There's a Whole World Of Other Screens Out There'

"Real" Examples:

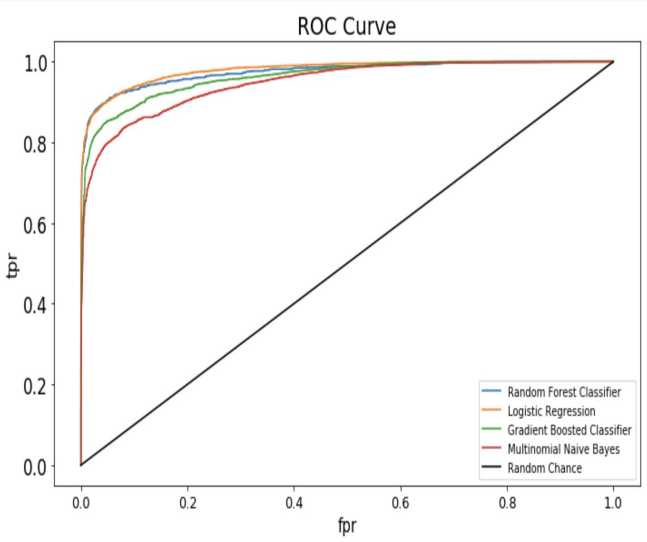
Reuters: 'The chairman of the Senate Foreign Relations Committee said on Wednesday his panel was likely to strip out language imposing sanctions against North Korea…'

NYT: 'Google is a top target for European regulators and privacy watchdogs, who openly fear and distrust its dominance. The American tech giant's…'

MODEL TESTING

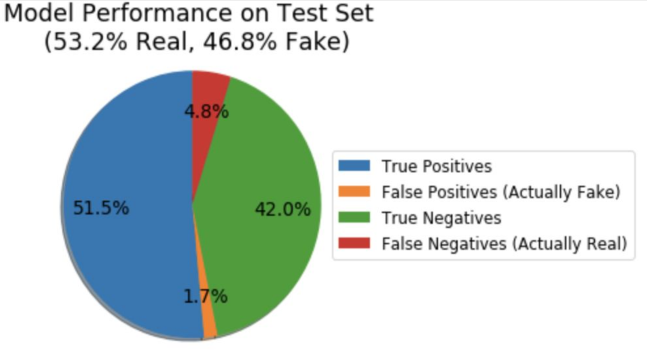
CROSS VALIDATION ON 80% OF DATA:

(Baseline 53%)	Logistic Regression	Random Forest Classifier	Gradient Boosted Classifier	Naive Bayes Classifier	Bidirectional LSTM
Accuracy	92.7%	92.8%	89.6%	87.5%	93.1%
Precision	94.5%	95.9%	90.1%	87.8%	92.2%
Recall	89.6%	89.0%	87.3%	85.0%	93.2%



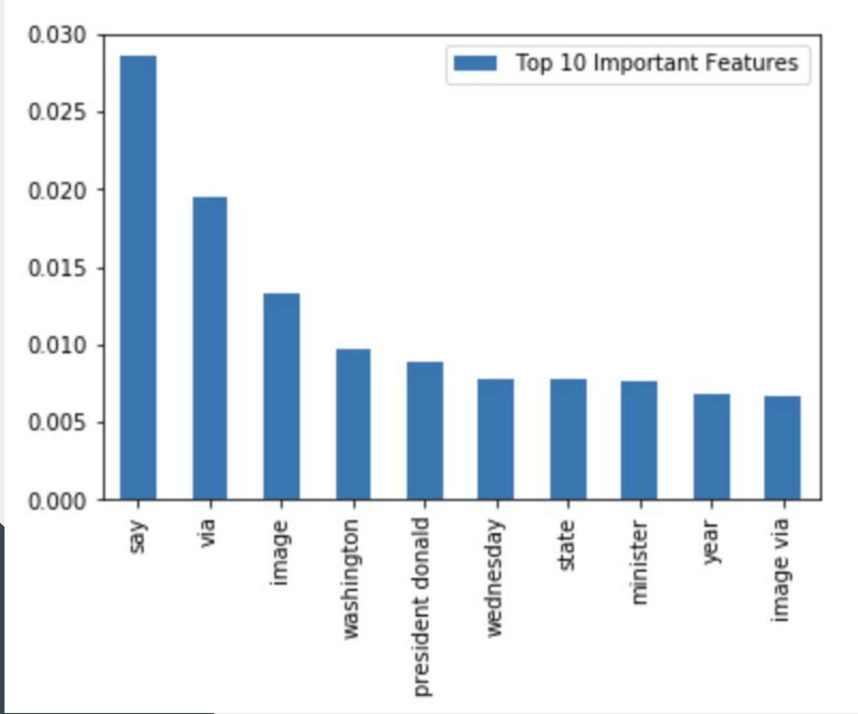
FINAL 20% HOLDOUT SET (RFC):

Accuracy	Precision	Recall
93.6%	96.9%	89.1%

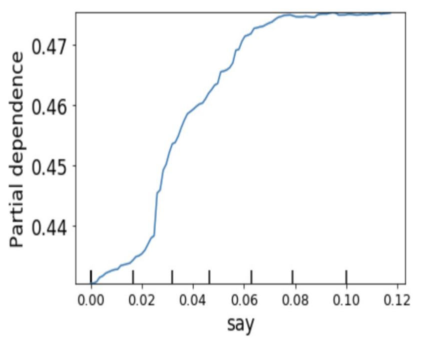


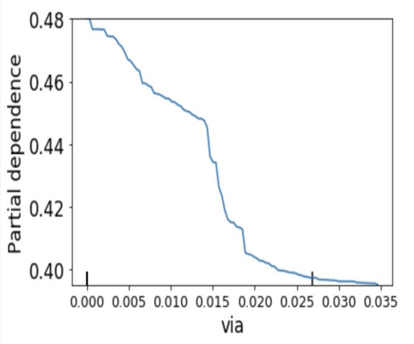
USING RFC TO FIND IMPORTANT WORDS

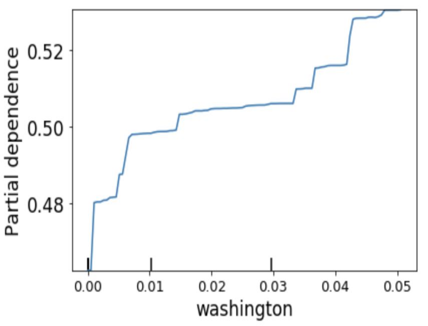
TOP INFORMATION GAIN WORDS:

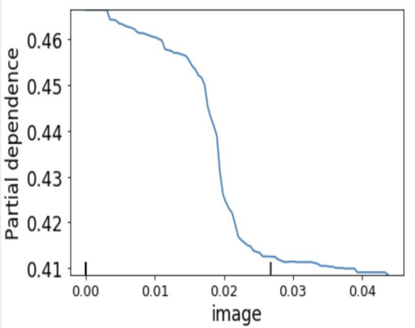


PARTIAL DEPENDENCE PLOTS:









CONCLUSION

This model does NOT do any sort of fact-check searching, only vocabulary indication

The words associated with higher weights assigned by the RFC have the largest effect on classification outcome.

FUTURE DIRECTION

Use BERT to give even longer range “context” to the words

Gather metadata about the articles to simulate actual fact-checking

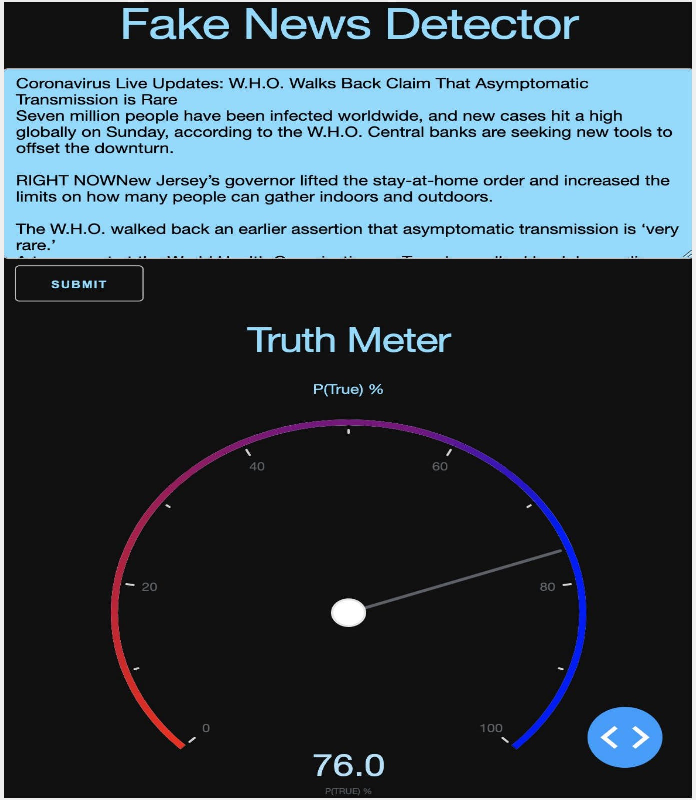
MY APPLICATION

[Link to Application]

Paste your article here! →

Submit →

Probability of being true (0-100%) →



THANK YOU!

AND STAY CLASSY, LA

Email: benelliot16@gmail.com
GitHub: github.com/benkarlsberg
LinkedIn: linkedin.com/in/benjaminkarlsberg/

APPENDIX

[Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism]

[Big Data and quality data for fake news and misinformation detection]

[IEEE-Dataport Fake News Data]

[Fake and Real News Dataset - Kaggle]

COMPARING TN TO FP

TRUE NEGATIVE (FAKE) EXAMPLE:

'Hillary Clinton is a Planned Parenthood's champion. She was greeted like a rock star at the Planned Parenthood Action Fund event in June, 2016. She mocked Donald Trump's defense of life, while highlighting her undying support for destroying the life of the most vulnerable. She's been bending over backwards to pander to Blacks for their votes, while trying desperately to gain the confidence of the Black Lives Matter movement.'

FALSE POSITIVE (ACTUALLY FAKE) EXAMPLE:

'Mexicans living abroad sent home a record amount of money in 2016, taking advantage of a strong U.S. labor market and a weakening Mexican peso amid worries about actions that the administration of U.S. President Donald Trump may take against immigrants or remittances. After a 25% jump in November, the month that Mr. Trump was elected, December remittances grew 6.2% from a year before to $2.34 billion, bringing the total for the year to $27 billion, the Bank of Mexico said Wednesday.'