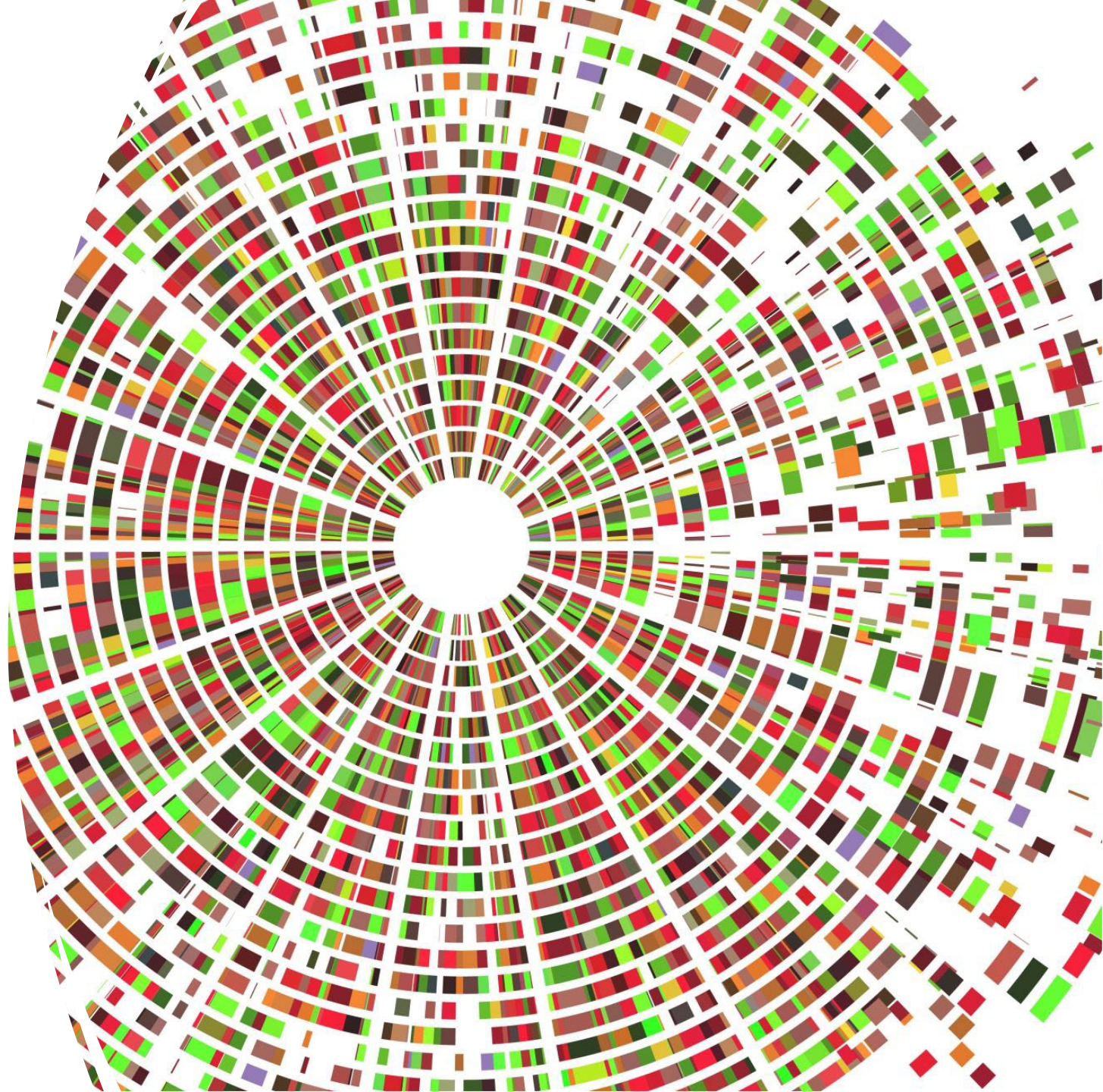


THE RISE OF FAKE NEWS IN THE POLITICAL SCENE

By Michael Mahoney





PRESENTATION OVERVIEW

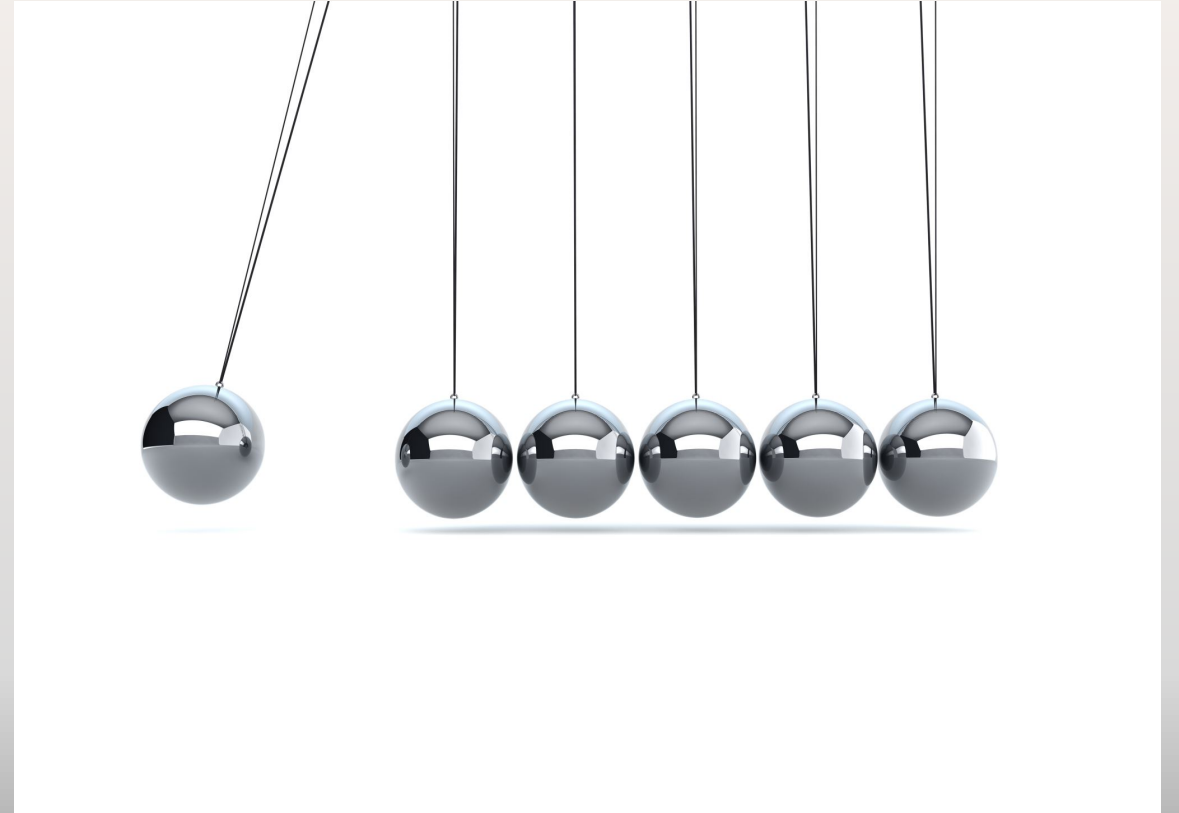
- Quick intro to the data
- Talk about misinformation and what we found
- Explore some recommendations
- Talk about the future



WHO IS THIS PRESENTATION FOR

MISINFORMATION

- The Cost 79 Billion
 - 37 billion in losses on the stock market
 - 17 billion in losses annually from false/misleading financial information
 - 9.54 billion spent on reputation management for false information about companies.
 - 3 billion has been already spent to combat fake news.
 - 400 million spent by US public institutions defending against misinformation.
 - 250 million spent on brand safety measure.

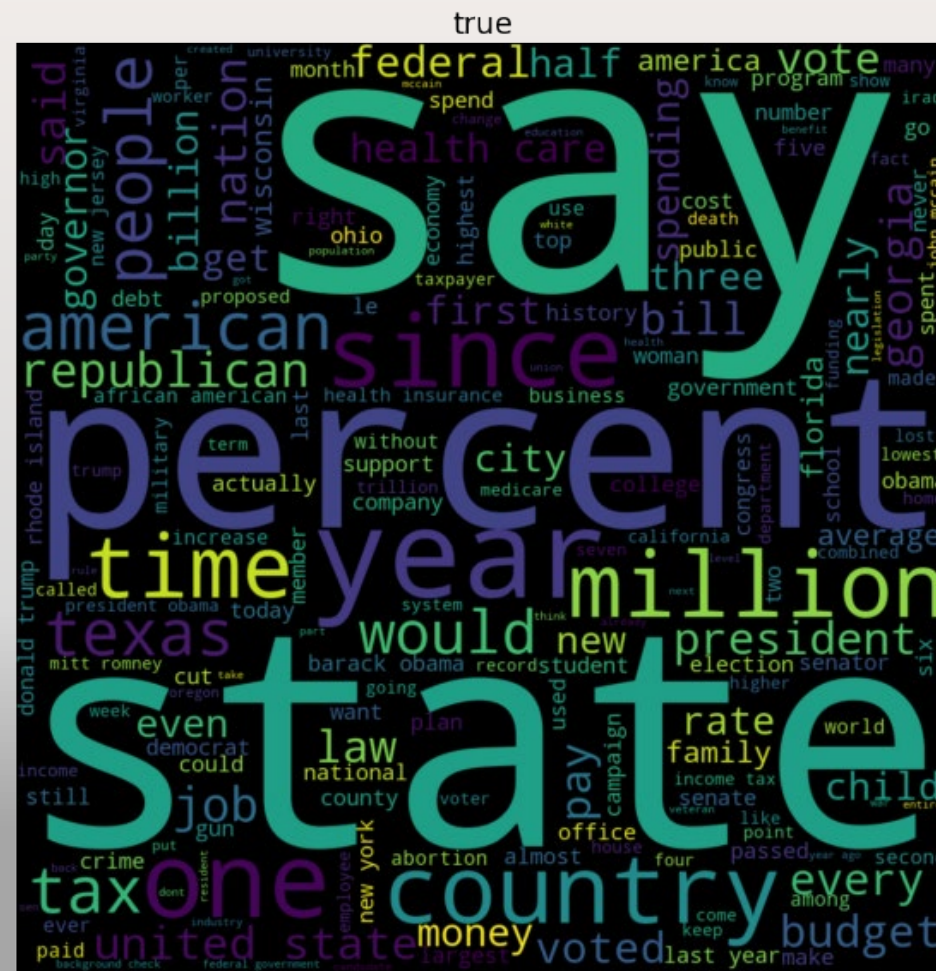
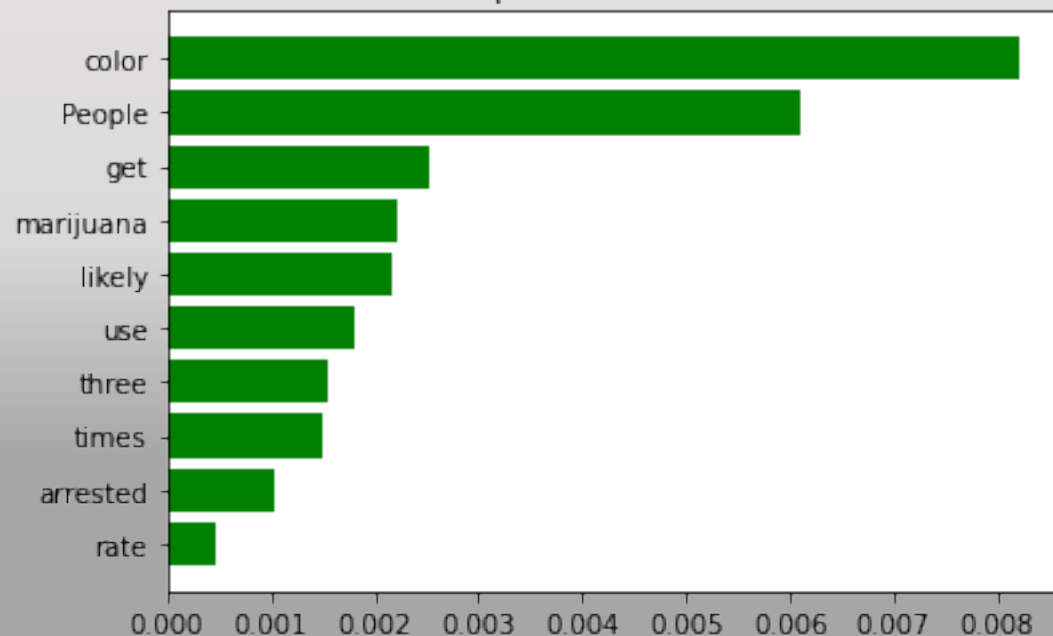


TRUE

TRUE QUOTE

People of color and those of not, they use marijuana at the same rate. People of color are three times more likely to get arrested and convicted.

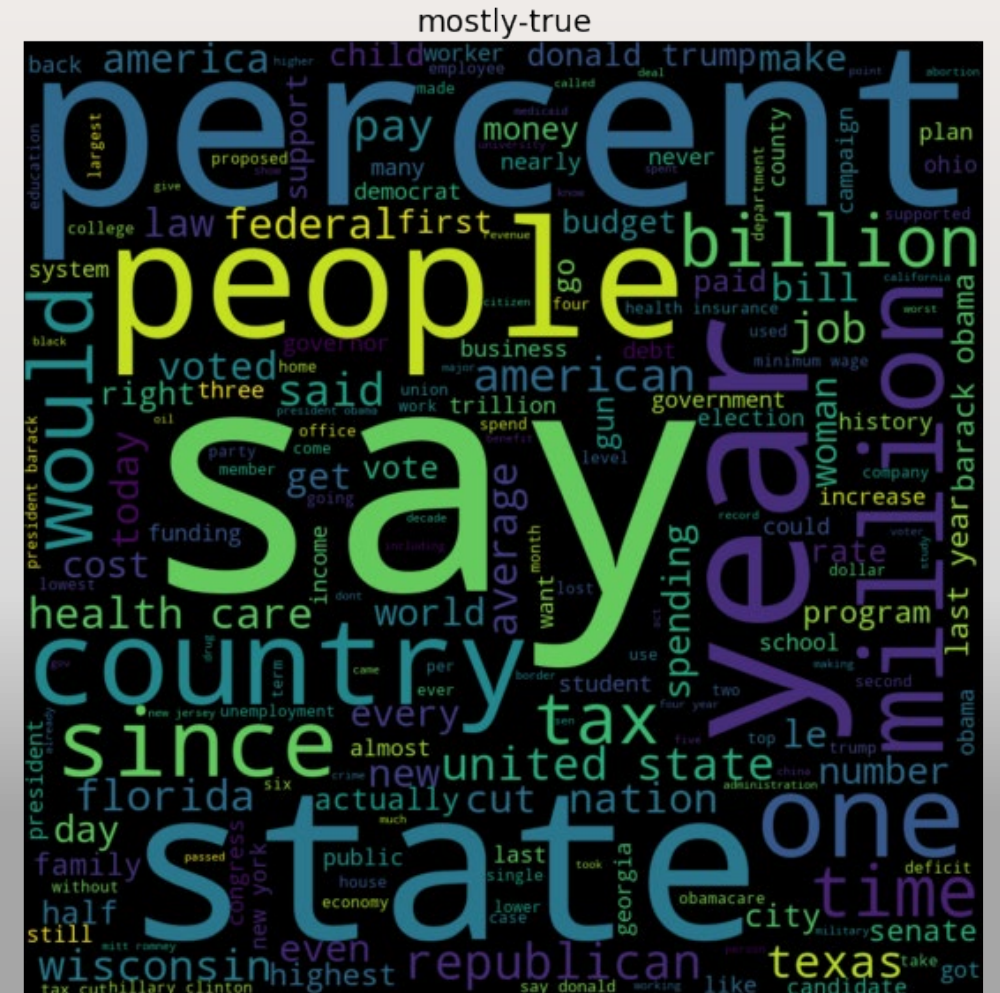
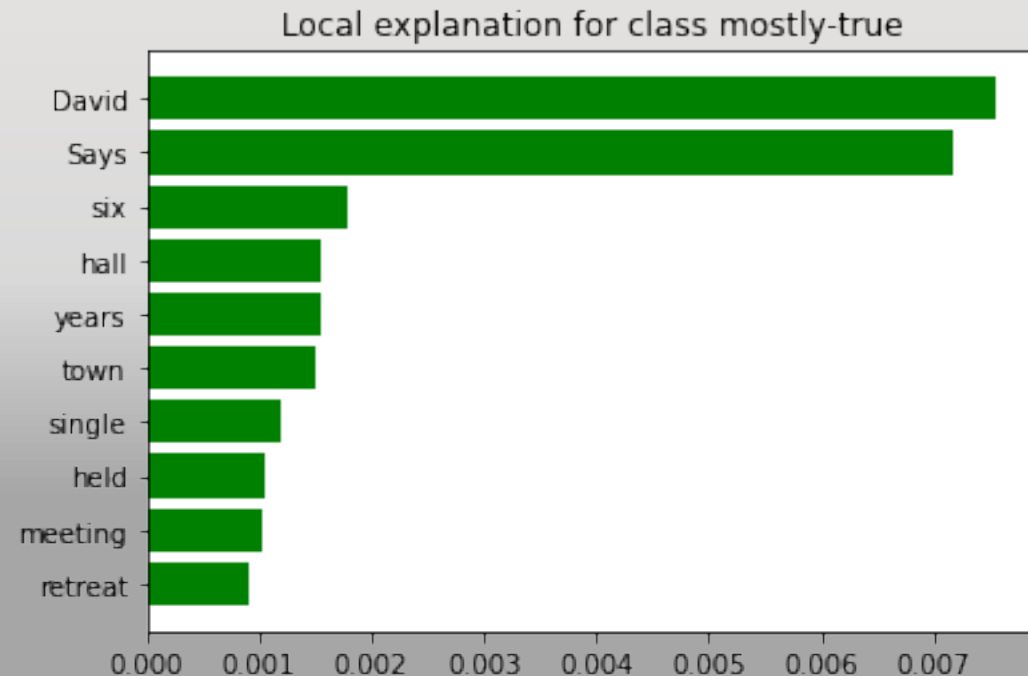
Local explanation for class true



MOSTLY-TRUE

MOSTLY-TRUE QUOTE

Says David Perdue hasn't held a single public town hall meeting in six years and sells four meetings a year and a retreat on a private island for a \$7,500 corporate PAC check.

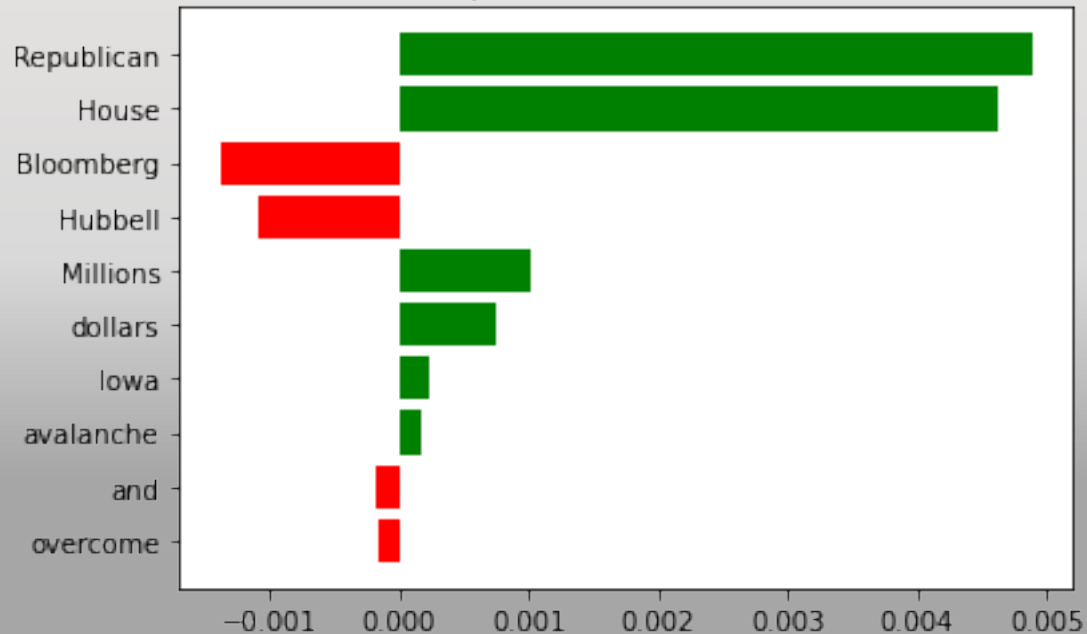


HALF-TRUE

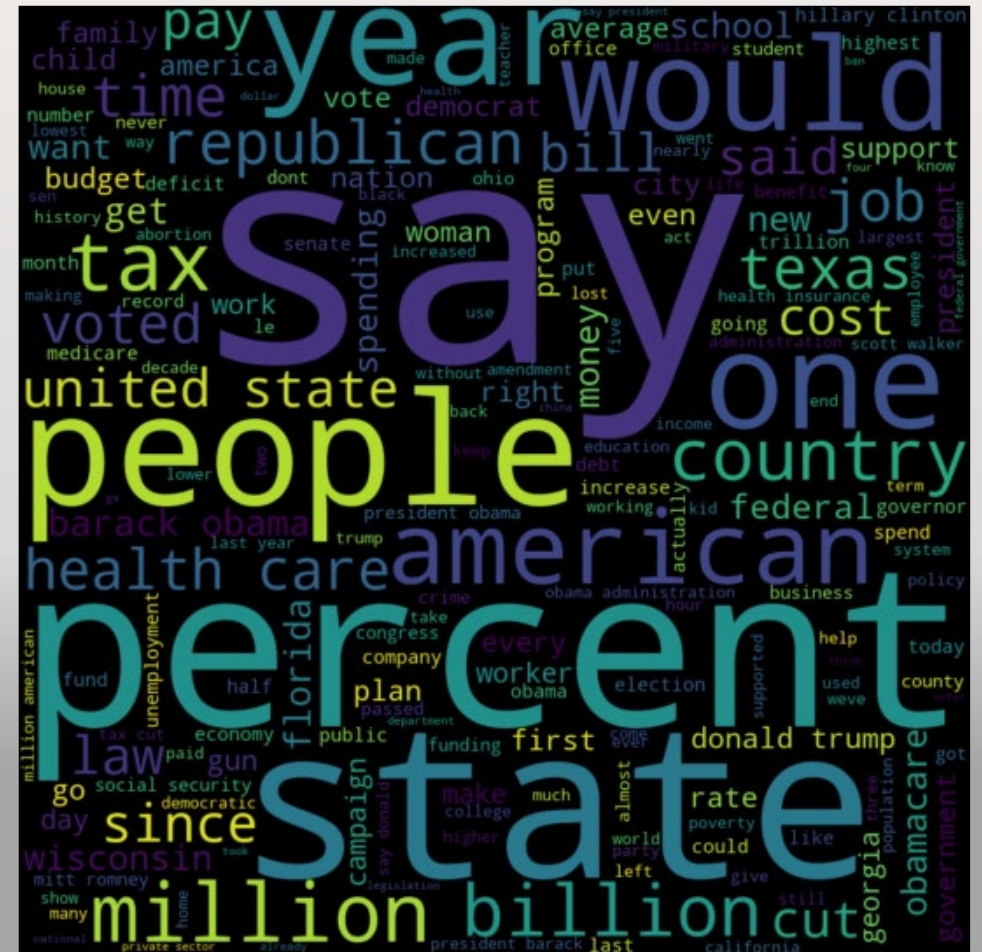
HALF-TRUE

Millions of Hubbell and Bloomberg dollars wasn't enough to overcome the Iowa House Republican avalanche of success.'

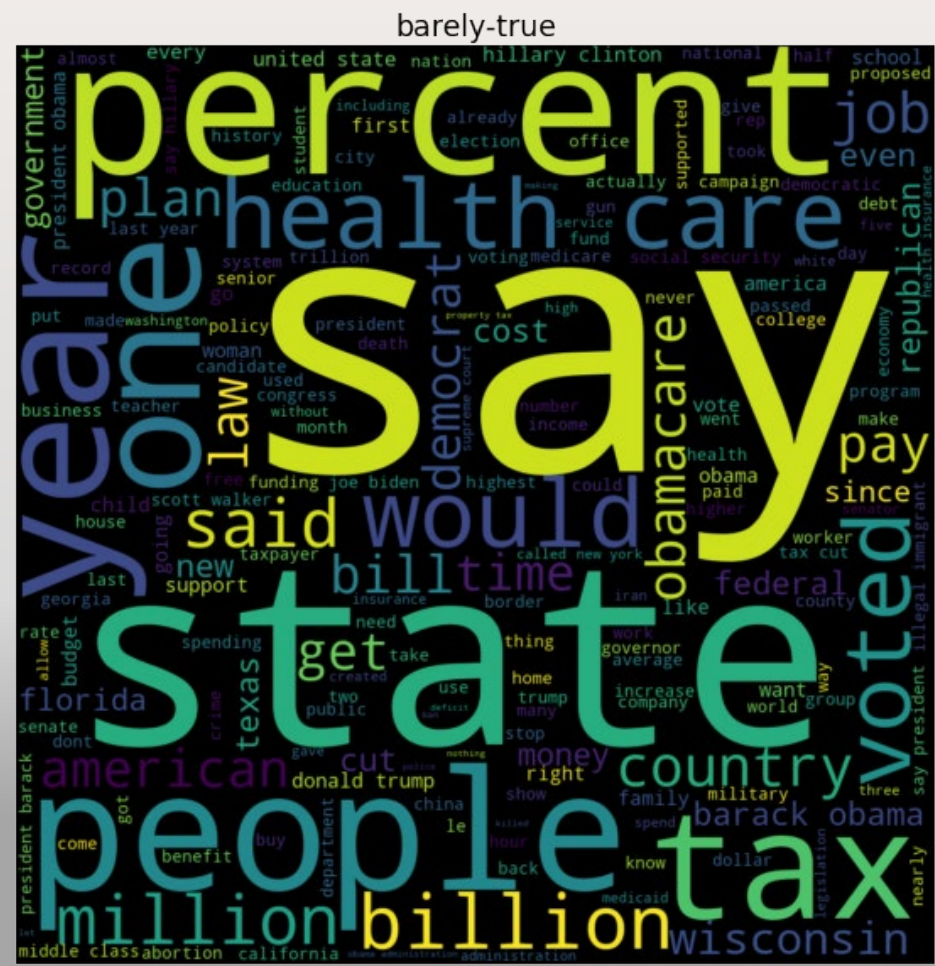
Local explanation for class half-true



half-true



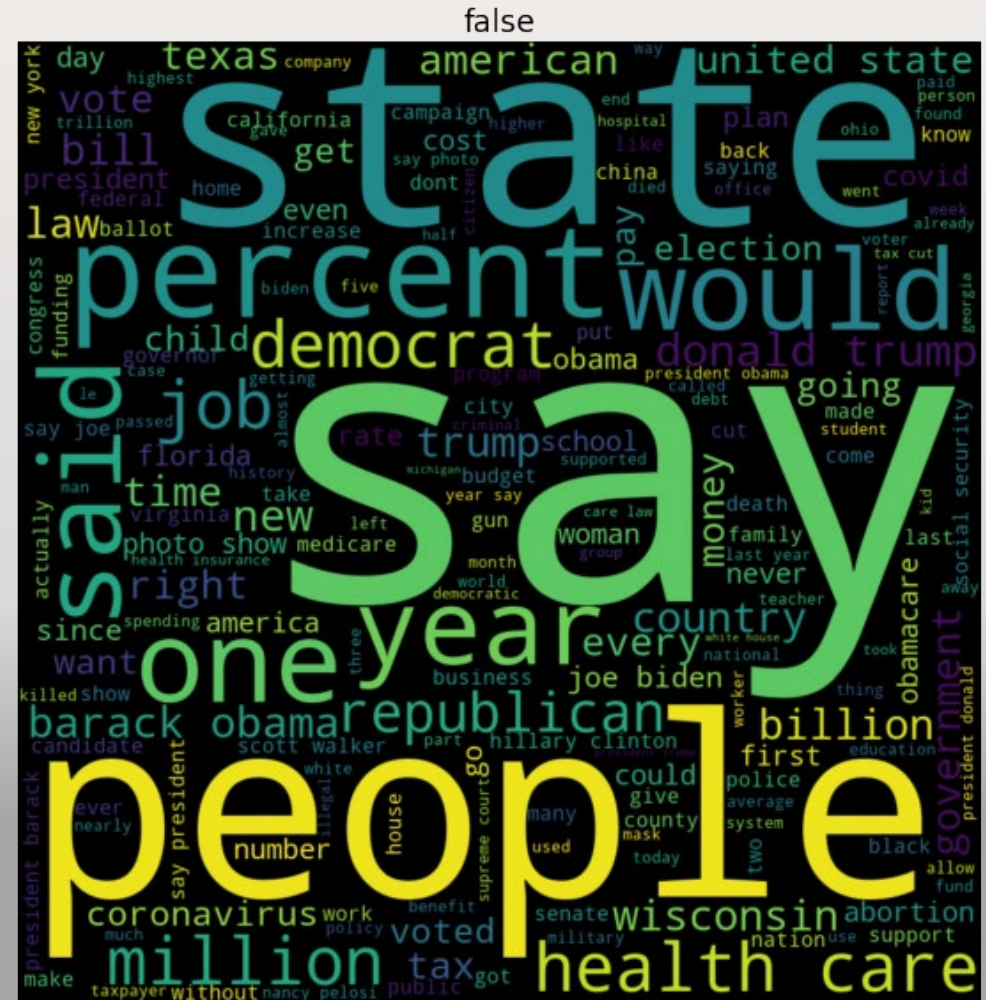
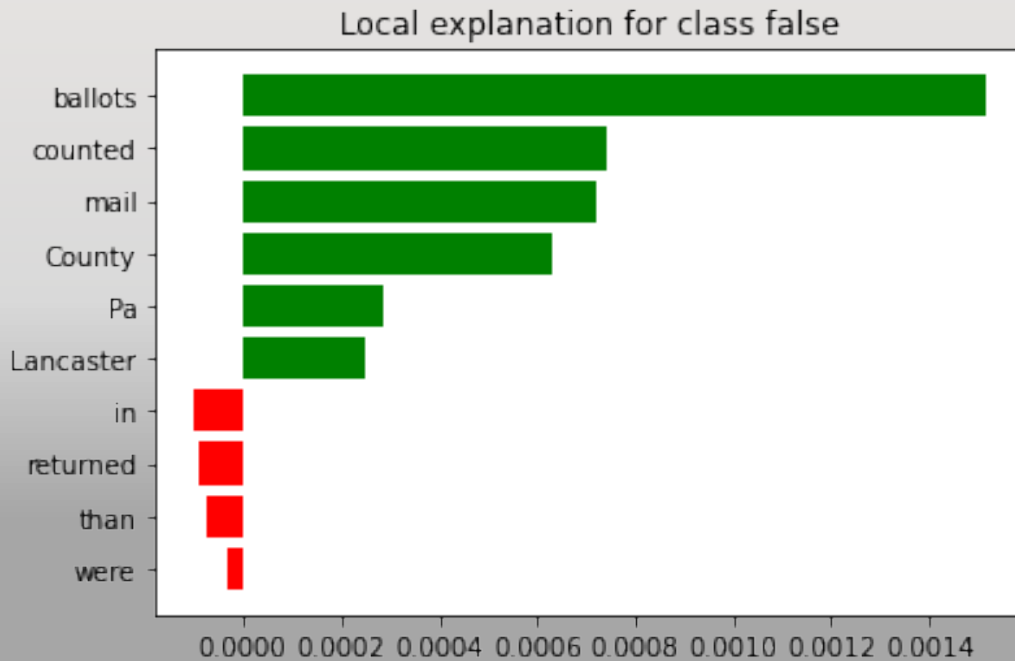
FDA announces 2 deaths of Pfizer vaccine trial participants from serious adverse events.



FALSE

FALSE QUOTE

More mail-in ballots were counted in Lancaster County, Pa., than were returned

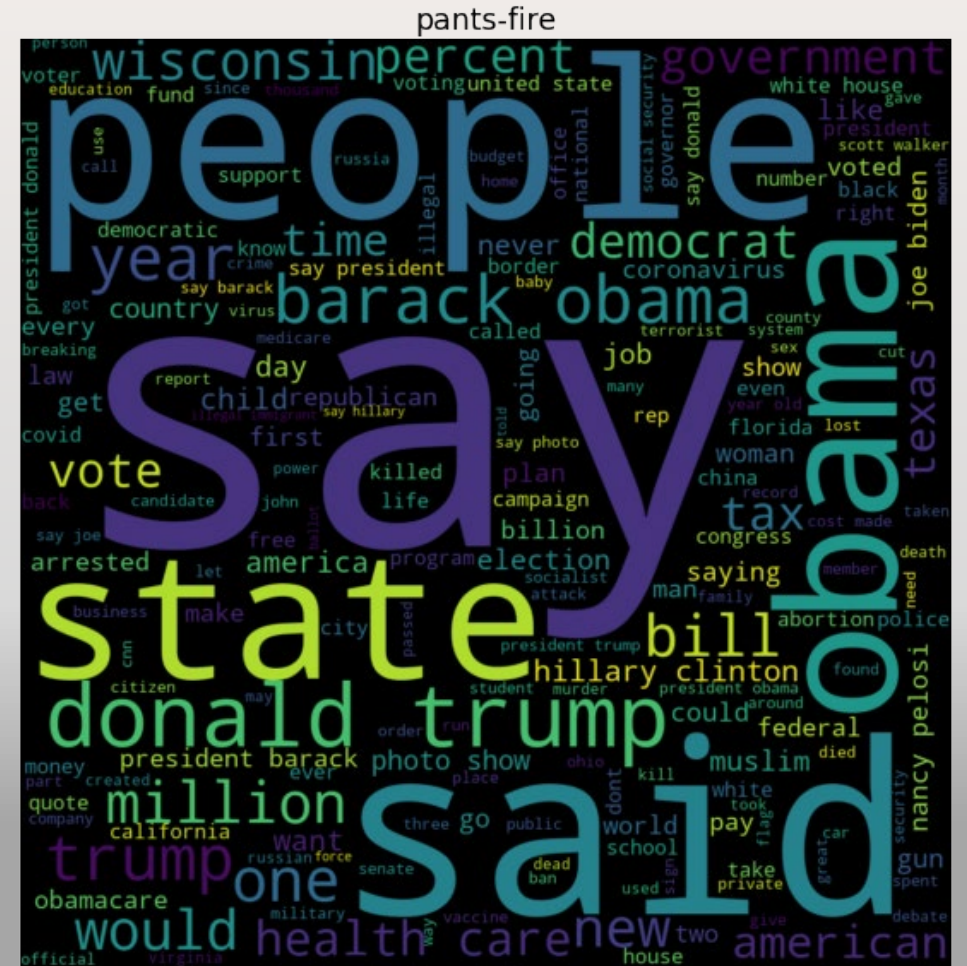
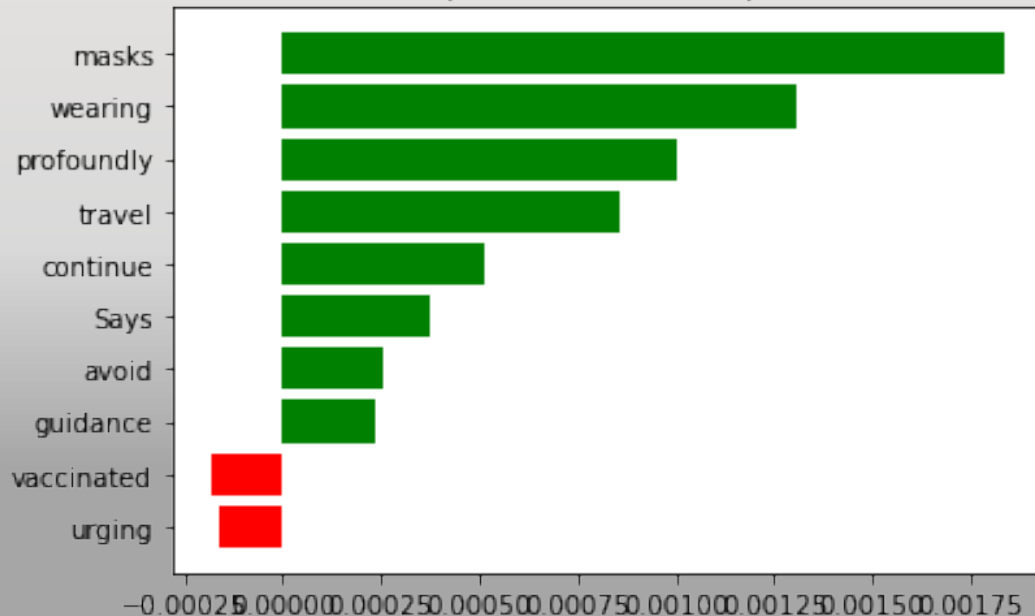


PANTS-FIRE

PANTS-FIRE QUOTE

Says that guidance urging vaccinated people to avoid travel and continue wearing masks is 'profoundly anti-science.'

Local explanation for class pants-fire

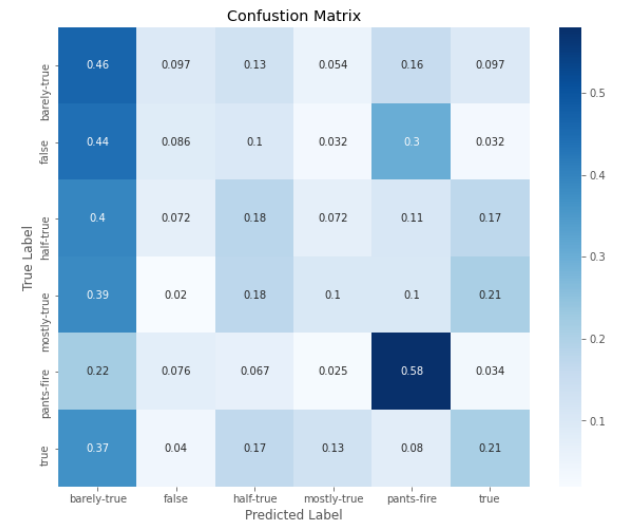
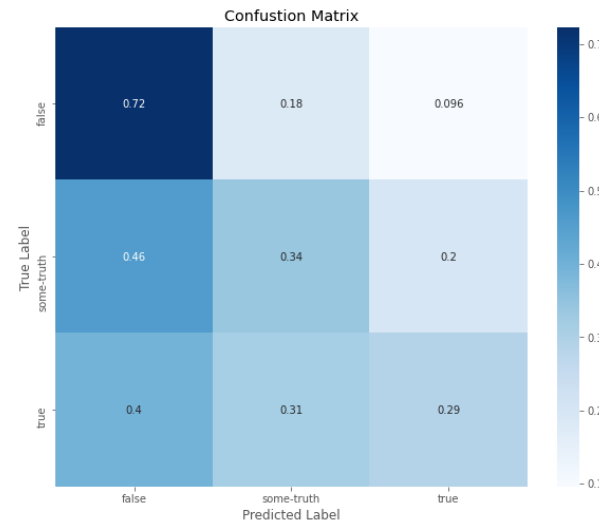
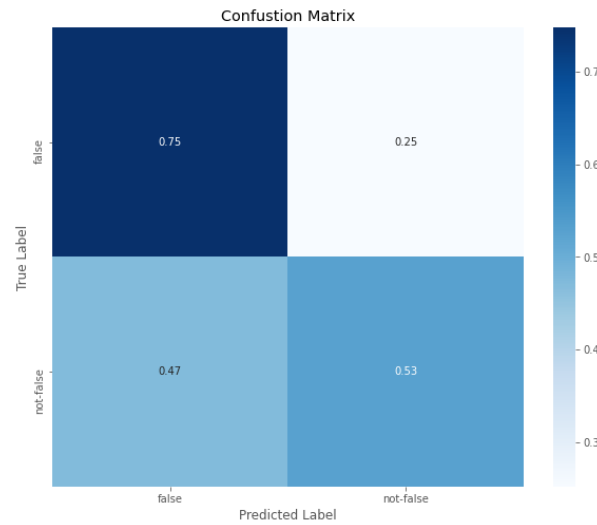
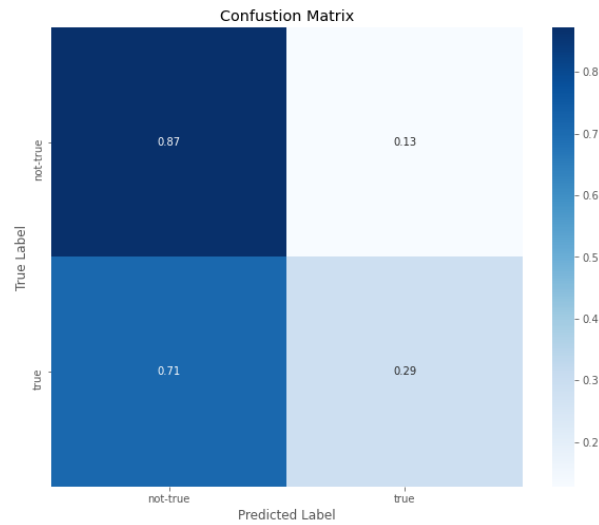




MISINFORMATION

- False is easyish
- True is less easy
- Anything else is hard

MODELS





WHAT TO DO ABOUT IT?

Resource Dedication

- For those wishing to devote resources to fight misinformation; virtually all clusters of the false quotes center around presidential elections and candidates. Resources should be ramped up and deployed at times of heightened contention during presidential elections and primaries.





Politics

- For politicians, most misinformation is based around people and not issues (with the notable exception of health care). If you wish to advance legislation, decoupling the issue from individuals will be the best way to limit misinformation's impact.

Perpetuating Information

- For Journalists, bloggers and people at large: Email chains, social media posts and viral content should never be used as a source for content.



FUTURE WORK

- I would like to bring in some more quotes.
- The second modeling goal is to implement a version of a BERT transformer.
- As for the data exploration; I would like to make more use of the various NLTK apis for text engineering.
- I would like to make the data more interactive on the app such that users can select specific quotes and interact with them through the various text models.

REFERENCES

- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is All you Need. NIPS 2017: 5998-6008
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT (1) 2019: 4171-4186
- Keras Developers: Keras API Docs: <https://keras.io/api/>
- Ployly Developers: Plotly API Docs: <https://dash.plotly.com/>
- Scikit Learn Developers: Scikit Learn API Docs: <https://scikit-learn.org/stable/modules/classes.html>
- Pandas Developers: Pandas API Docs: <https://github.com/pandas-dev/pandas>
- Jacob Kessler: ScatterText: <https://github.com/JasonKessler/scattertext>
- Marco Tulio Correia Ribeiro: Lime: <https://github.com/marcotcr/lime>
- Radim Rehrek, Petr Sojka: Software Framework for Topic Modelling with Large Corpora: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks: p. 45 - 50 : 2010: ELRA: Valletta, Malta
- Andreas Mueller: word-cloud: https://github.com/amueller/word_cloud

THANK YOU

