metis nyc ds9
ting neo
project fletcher
november 11, 2016

topic modeling online petitions

change.org

- "The World's Platform for Change" (a.k.a, the go-to site for online petitions and web-uprisings)

President Obama signs Bill of Rights for Sexual Assault Survivors

140,739 supporters

Keep Mr. Zlolek as a Full-Time Choir Director

139 supporters

Electoral College: Make Hillary Clinton President on December 19

1,981,441 supporters

business scenario

As a hypothetical data scientist at change.org, I might be...

- ☐ Topic modeling to improve business intelligence
    - ■ What topics do people care about?
    - ■ Which topics are the most successful in producing change?

- ☐ Developing functionality for users
    - ■ Petition dashboard
    - ■ Display topic importances
    - ■ Petition recommendations

data and methods

50,000+ petitions > 30 topics

Data Collection & Storage

Scrapy

Selenium

mongoDB

Features

- ☐ Petition text
- ☐ Comment text
- ☐ Petition attributes (e.g., # supporters, goal, tags)

Modeling

Custom tokenizer

Sklearn

Techniques

- ☐ TF-IDF Vectorizer
- ☐ NMF
- ☐ Cosine similarity
- ☐ Count Vectorizer, LDA, K-Means, pipelines
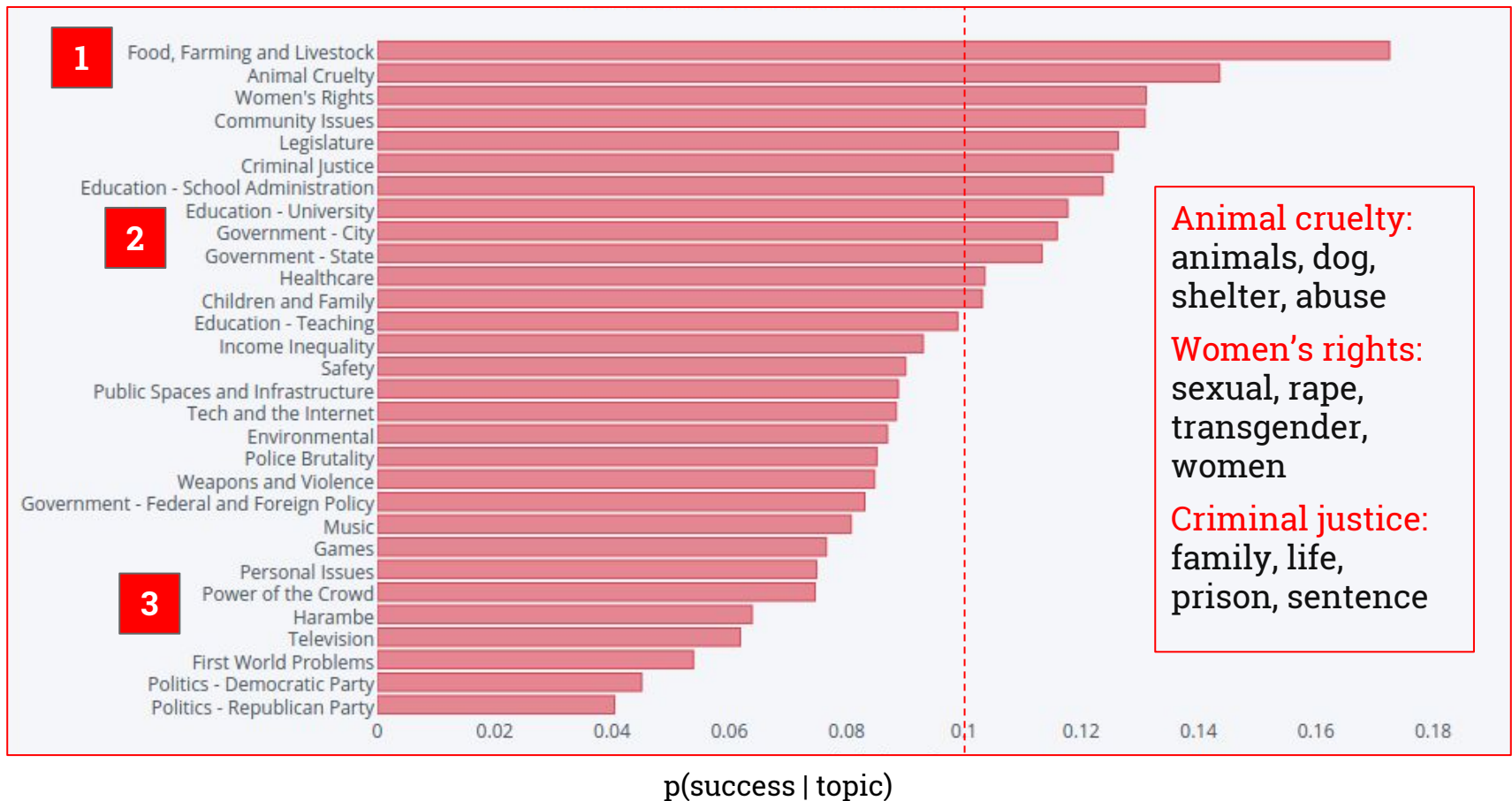
Implementation

Plotly

Flask

Deliverables

- ☐ Topic insights
- ☐ Petition dashboard

topics and conclusions

1. Petitions related to animal welfare and cruelty were most likely to succeed

2. Multiple topics appeared for both education and government, with layered nuances

3. Yes, Harambe and First World Problems are legitimate topics

Animal cruelty: animals, dog, shelter, abuse

Women's rights: sexual, rape, transgender, women

Criminal justice: family, life, prison, sentence

p(success | topic)

dashboard demo

future work

- Extend classification pipeline and build out dashboard to be able to provide probabilities of petition success

Change Gorilla Glass to Harambe Glass
6 supporters

Put Harambe on the Dollar Bill
18,329 supporters

Make Harambe a Pokemon
121,957 supporters

thanks!

ting neo

you can find me at

http://github.com/neokt

neo.kaiting@gmail.com

topic modeling
online petitions