

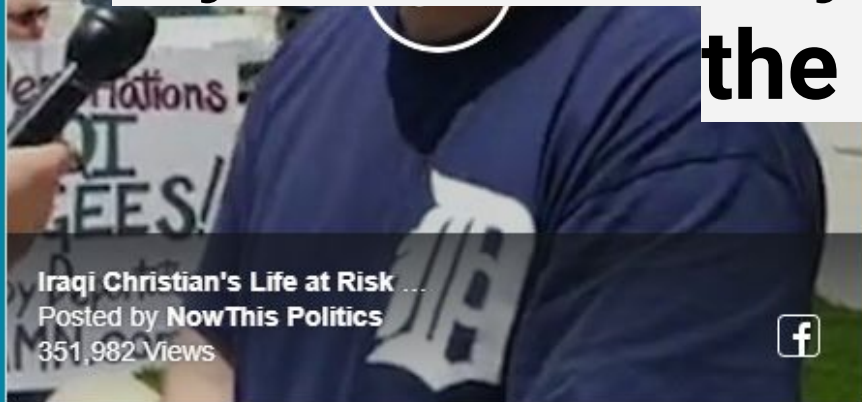
# Recommender systems are Breaking Democracy

Envisioning an alternative kind of  
recommender





**My internet and your internet are not the same**



**Iraqi Christian's Life at Risk ...**

Posted by **NowThis Politics**

351,982 Views



'A return to Iraq would be a death sentence.' — This man has lived in the U.S since he was 3 years old, but the Trump admin could put his life at risk by deporting him to Iraq



**Conservative Post** ✓

about 3 weeks ago



Should we be required to change our way of life to not offend immigrants?

I thought it was the other way around...



CONSERVATIVEPOST.COM

**Here's What Happened When a Muslim Immigrant Co...**

Muslim immigrants believe they can manipulate our legal system ...



89



350



127

83% of us are online daily  
28% say are online 'constantly'  
3.5 hours/day average online

Much of that online experience is curated by

## **Recommendation Systems**

serving you content you probably already agree with.

Recommender systems watch what you click and push content you are already likely to agree with

**neglecting content  
from opposing  
viewpoints**



LIBERAL ⓘ

SHOWING POSTS ABOUT:

"PRESIDENT TRUMP"

CONSERVATIVE ⓘ

490 108 418



MSNBC ✓

about 3 weeks ago



Reverend Al Sharpton : "No one wants to play with you, Mr. President, outside of your shrinking base and maybe a few dictators. Before you run crying to Fox News, remember their polling has you losi

Which leads us back to this.

We are on the same planet but living in two different worlds.



about 3 weeks ago

Under President Trump, African American and Hispanic unemployment has hit record lows, resulting in the greatest jobs outlook for black Americans and Latinos in the history of the United States.

Or, as Cher puts it:



Cher: Trump 'More Than a White Supremacist'

204

993

192

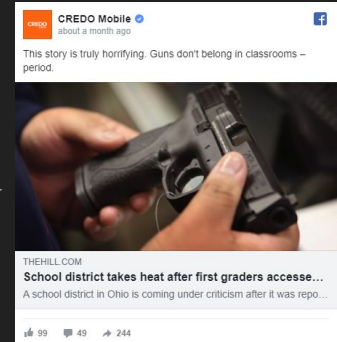
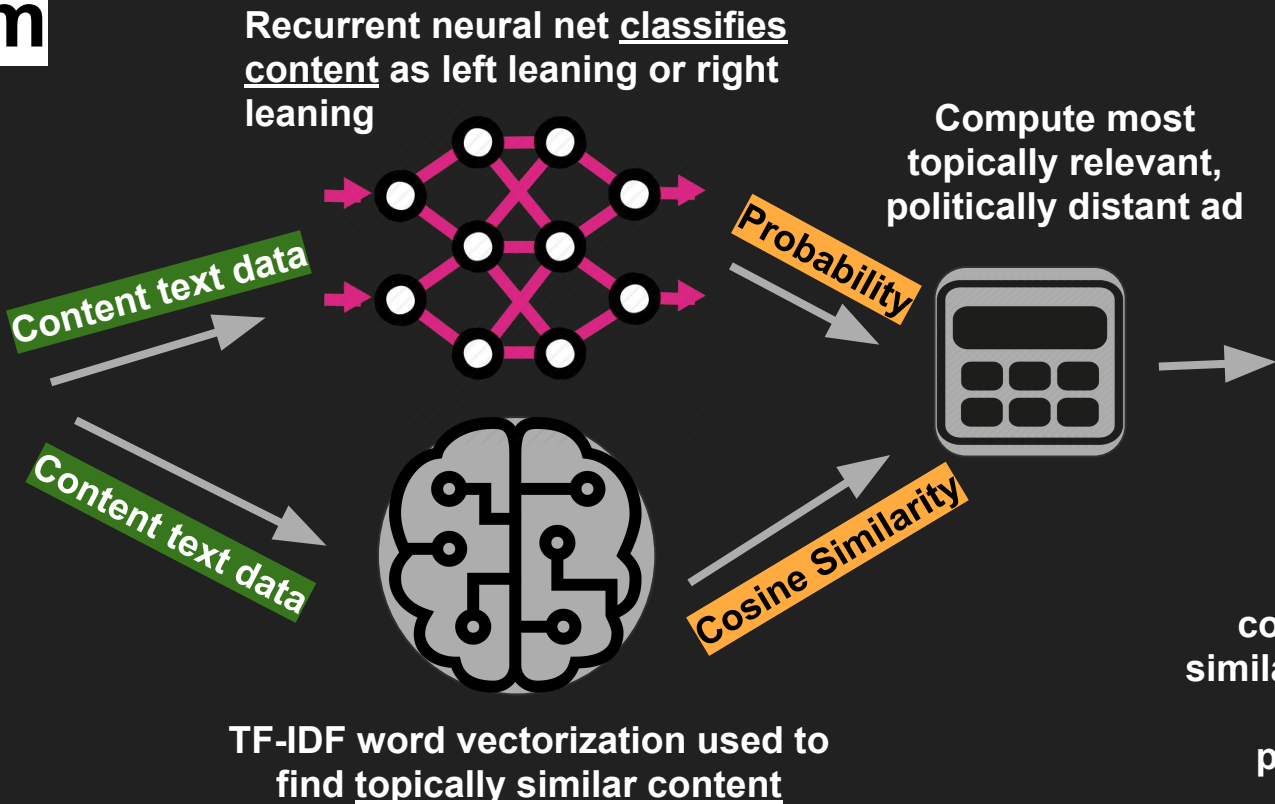
MSNBC.COM

Memo to Trump: 'No one wants to play with you'

Imagining a recommender  
system that **fights filter**  
**bubbles and echo chambers**  
what would that look like?

# A non-partisan recommender that serves topically relevant content from the opposite political spectrum

User clicks content



Recommended content is topically similar to original, but from opposite political spectrum

# Data



## FACEBOOK DATA

I used political ad data from facebook's graph API. Free. Available.

Corpus of mostly partisan content spanning many topics to train from.

# FACEBOOK DATA

**800,000 political ads**

**Text from the ads** - used to classify liberal/conservative

**Link to the ad** - used to display recommended ad

**Who paid for the ad**

Facebook does not allow raw data to be shared so data will not be posted

# Data Wrangling

**Classifiers need labeled examples to learn from**

**Labeling 30,000 ads liberal/conservative ads by hand not  
feasible**

**Who paid for the ad used as a proxy for left right**

**Example Labeled Left**

**Senate Majority  
PAC**

**DCCC**

**Example Labeled Right**

**Restoration PAC**

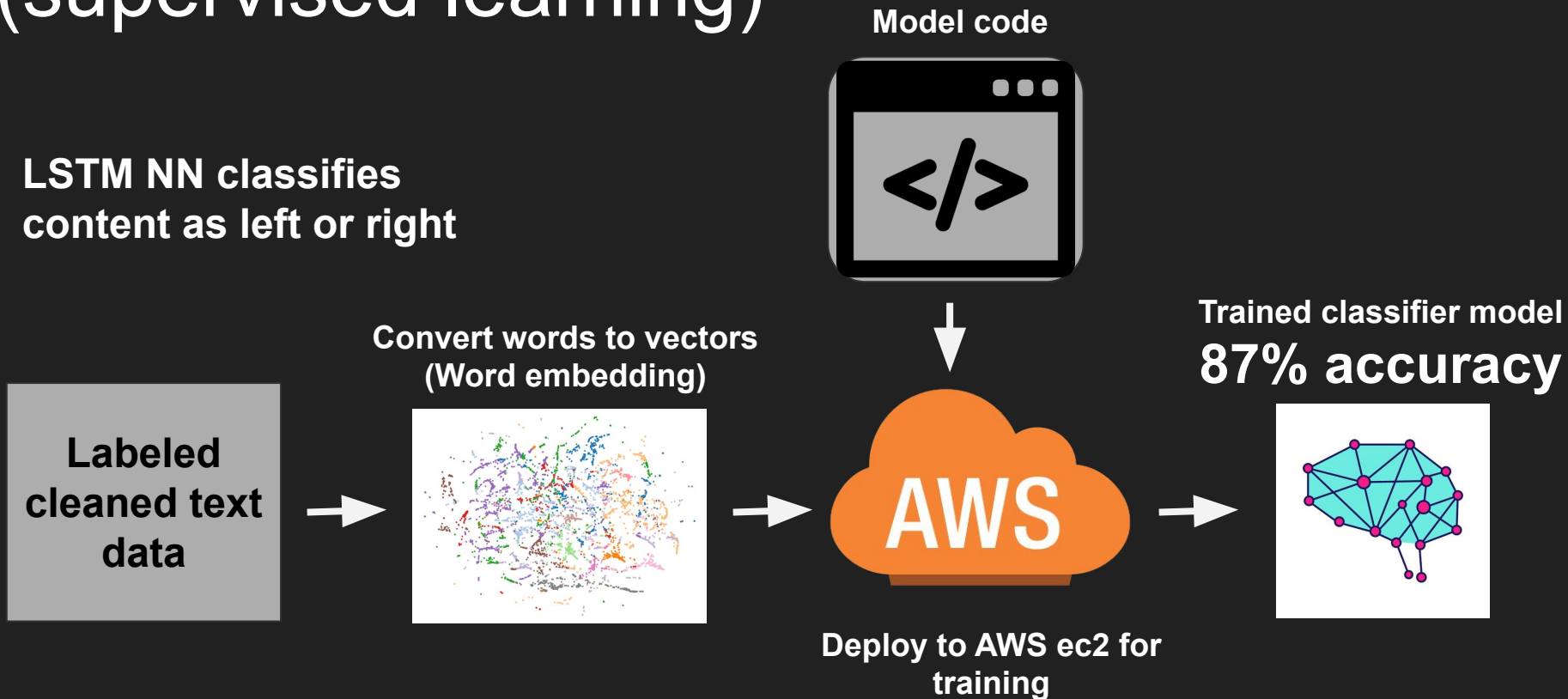
**Heritage  
Foundation**

# Different models need different data prep approaches

	Article similarity model TF-IDF/Cosine-sim	Partisan classification model LSTM NN
Remove special chars *^&/@\$% etc.		
Replace emoji with word representation		
Stem words to 'root word' ie "running" to "run"		
Convert sentences to TF-IDF vectors		
Convert words to word embedding vectors		

# Classification Model Training (supervised learning)

LSTM NN classifies  
content as left or right



# Topicality model TF-IDF + Cosine Similarity (unsupervised learning)

Stemmed  
cleaned text  
data



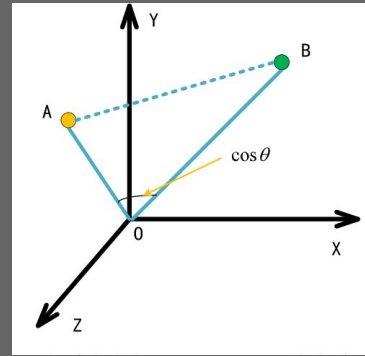
$$w_{x,y} = \text{tf}_{x,y} \times \log\left(\frac{N}{\text{df}_x}\right)$$

**TF-IDF**

Term  $x$  within document  $y$

$\text{tf}_{x,y}$  = frequency of  $x$  in  $y$   
 $\text{df}_x$  = number of documents containing  $x$   
 $N$  = total number of documents

Apply TF-IDF to turn  
documents into vectors



Cosine of angle between two  
document vectors measures  
their similarity



List of  
document  
similarities  
ranked by  
similarity to  
clicked  
content

# Combining classification and topicality

NN's don't always get things right.

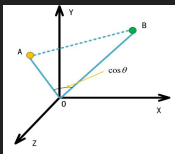
Combining the outputs of the two models in below reduces chance of serving recommending misclassified ads

Seeking left wing

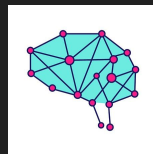
$$\textit{ContentSimilarity} \times \textit{ClassificationCertainty}^{25}$$

Seeking right wing

$$\textit{ContentSimilarity} \times (1 - \textit{ClassificationCertainty})^{25}$$



**Cosine sim topic model  
output between -1,1**




**LSTM classifier  
NN output between 0,1**

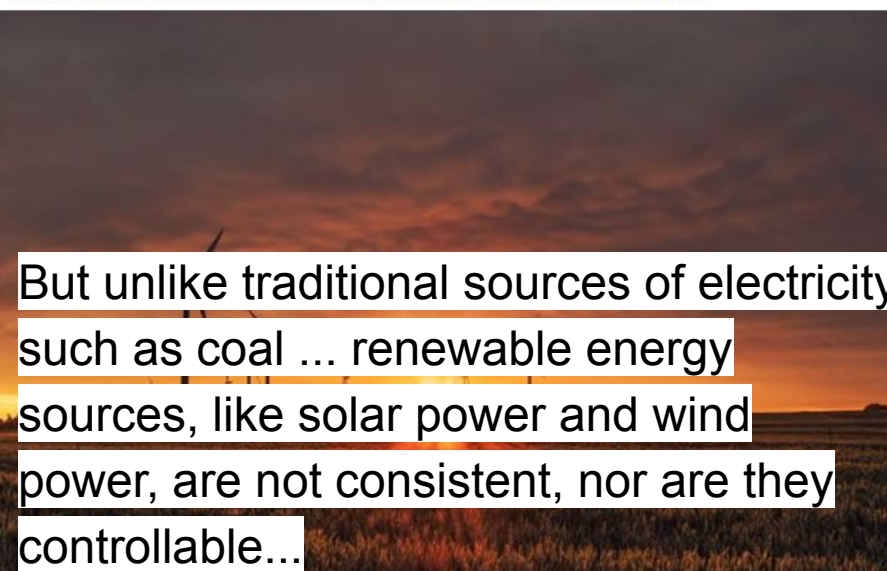


# Pro Fossil Fuel

## Clicked ad

 **Michigan Capitol Confidential**  
Sponsored • Paid for by The Mackinac Center for Public Policy

When you flip a light switch on, you expect an instant result. But unlike traditional sources of electricity such as coal, natural gas and nuclear energy, renewable energy sources, like solar power and wind power, are not consistent, nor are they controllable.



But unlike traditional sources of electricity such as coal ... renewable energy sources, like solar power and wind power, are not consistent, nor are they controllable...

# Anti Fossil Fuel

## Recommended ad

 **Suzan DelBene**  
Sponsored • Paid for by DelBene for Congress

Donald Trump's new proposal will weaken requirements on power companies to control their greenhouse gas emissions. It rolls back the progress the EPA made under President Obama to push electric companies to use newer, cleaner sources of energy, like natural gas and renewables, and to move away from dirty, outdated methods like burning coal.

Trump's so-called Affordable Clean Energy rule is neither of those things.



Donald Trump's new proposal will weaken requirements on power companies to control their greenhouse gas emissions...

# Attack ad against Joaquin Castro

## Clicked ad

**RUM** Trump Victory Fund  
Sponsored • Paid for by Trump Victory Fund

Attacking the donors of the Trump campaign is not only wrong it's dangerous. Tell Democrats like Joaquin Castro and socialist Rashida Tlaib to stop endangering hardworking, patriotic Americans.

Ad removed



...Tell Democrats like Joaquin Castro and socialist Rashida Tlaib to stop endangering hardworking, patriotic Americans

# Support ad for Joaquin Castro

## Recommended ad

**Texas Democratic Party**  
Sponsored • Paid for by the Texas Democratic Party

Congressman Joaquin Castro is rejecting Trump's FAKE national emergency and standing up to Trump's stupid border wall. Add your name to thank Rep. Castro for his fight to stop Trump from stealing Texas land!  
<https://bit.ly/2JHfdpN>

...Add your name to thank Rep. Castro for his fight to stop Trump from stealing Texas land


REJECTING TRUMP'S FAKE EMERGENCY

ADD MY NAME



# Democratic messaging related to trump supreme court nominees

## Clicked ad

 End Citizens United  
Sponsored • Paid for by End Citizens United

Beto O'Rourke is SO CLOSE to DESTROYING Ted Cruz!! Democrats NEED him to win to take back the Senate & SAVE the Supreme Court from Trump's extreme nominee. If we lose this race, we'll LOSE the Supreme Court for decades to come!! PLEASE rush \$15 to flip Texas blue and SAVE the Supreme Court!!

**IF WE DEFEAT TED CRUZ,  
DEMOCRATS CAN SAVE THE**

**Beto O'Rourke is SO CLOSE to DESTROYING Ted Cruz!! Democrats NEED him to win to take back the Senate & SAVE the Supreme Court from Trump's extreme nominee. If we lose this race, we'll LOSE the Supreme Court for decades to come!! PLEASE rush \$15 to flip Texas blue and SAVE the Supreme Court!!**

[Donate Now](#)

# Republican messaging related to supreme court nominees

## Recommended ad

 Kevin McCarthy  
Sponsored • Paid for by Kevin McCarthy for Congress

**BREAKING: President Trump's Supreme Court nominee has been CONFIRMED! Show your support below >>>**

**KAVANAUGH CONFIRMED!**  
**SHOW YOUR SUPPORT**  
**ADD YOUR NAME**

**Donald Trump's new proposal will weaken requirements on power companies to control their greenhouse gas emissions...**

[Sign Up](#)

**BREAKING: Judge Kavanaugh confirmed!!! Add your name to show your support>>>**

# How we can improve performance

- 1. Ads are not just text, content in images too**
  - a. Add optical character recognition
- 2. Non political content in corpus (classic open set problem)**
  - a. Train model to recognize degree of partisanship
    - i. would need existing label dataset or crowd source labeling
- 3. Content with lower probability of correct left/right classification is less likely to be recommended to users in current state algorithm**
  - a. Need to improve LSTM performance to reduce reliance on content downgrading
  - b. experiment with more topic modeling methods to quantify imbalances, pull more topically relevant data to make up for imbalances and retrain
- 4. Right now is programmed to work with english content only**
  - a. Could be modified relatively easily to work with other common languages

# Contact

Joe McAllister

[Email](#) [Github](#) [Medium](#) [Linkedin](#)

# References

left/right wing graphics -> <http://graphics.wsj.com/blue-feed-red-feed/#/immigration>