Loaded Topics: Predicting bias through story selection and word choice

Nick Thomas

"I trust this site to tell the truth."

Determining Bias

- Hostile Media Effect

- Way to predict bias from data?

Media bias chart showing news sources categorized by political lean (Garbage Left, Hyper-Partisan Left, Leans Left, Neutral, Leans Right, Hyper-Partisan Right, Garbage Right) and by quality (High Quality, Mixed Quality).

GARBAGE LEFT (NOT WORTH IT)

HYPER-PARTISAN LEFT (TO CONFIRM YOUR BELIEFS)

LEANS LEFT (NOT HORRIBLE)

NEUTRAL (WHAT JOURNALISM SHOULD BE)

LEANS RIGHT (NOT HORRIBLE)

HYPER-PARTISAN RIGHT (TO CONFIRM YOUR BELIEFS)

GARBAGE RIGHT (NOT WORTH IT)

HIGH QUALITY (GOOD THREAD)

MIXED QUALITY (BETTER THAN NOTHING)

High Quality sources: ABC NEWS, NBC, npr, AP, C-SPAN, CATO INSTITUTE, THE DAILY BEAST, BROOKINGS, CBS NEWS, BBC NEWS, The Economist, Forbes, the weekly Standard, The Daily Dot, Bloomberg, REUTERS, WSJ, NATIONAL REVIEW, Mother Jones, FiveThirtyEight, MEDIA MATTERS, POLITICO, FT, reason, BLUE LIVES MATTER, The New York Times, THE TEXAS TRIBUNE, Rare, AMERICANS FOR PROSPERITY, BLACK LIVES MATTER, PBS NEWS HOUR, Examiner, The Washington Post, Atlantic, USA TODAY, SALON, Newsweek, theguardian, CNN, FOX NEWS

Mixed Quality sources: ALTERNET, Daily Mirror, MISES INSTITUTE, ALLEN B. WEST, DAILY KOS, MEDIAite, NY DAILY NEWS, THE HILL, DRUDGE REPORT, JEZEBEL, MoveOn.org, MSNBC, INDEPENDENT JOURNAL REVIEW, WND, ThinkProgress, ELITE DAILY, RT, RED ALERT POLITICS, THE DC, THE YOUNG TURKS, RedState, THE HUFFINGTON POST, VICE NEWS, NEW YORK POST, theblaze, Vox, BuzzFeed, TPM, BREITBART

Establishing Sources

- Web scraping using beautifulsoup and selenium
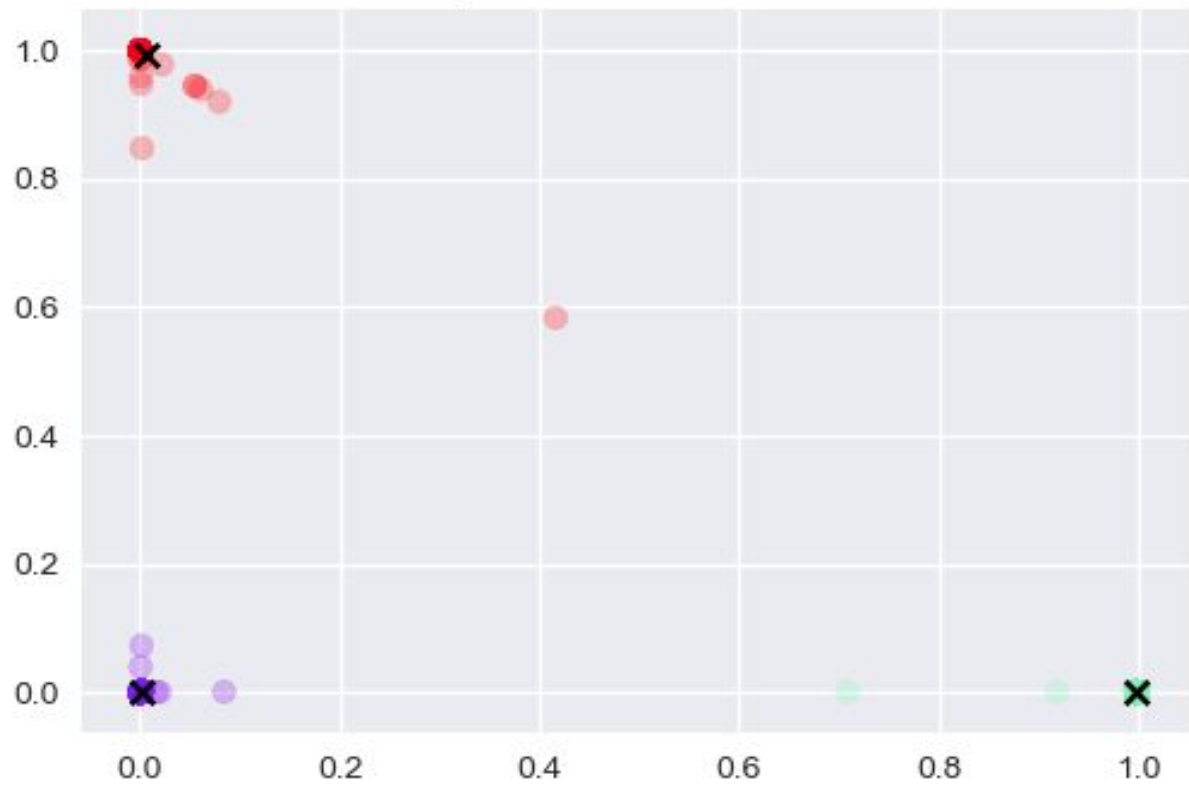
- Join articles in Mongodb

- Drop redundant articles

Topic Modeling
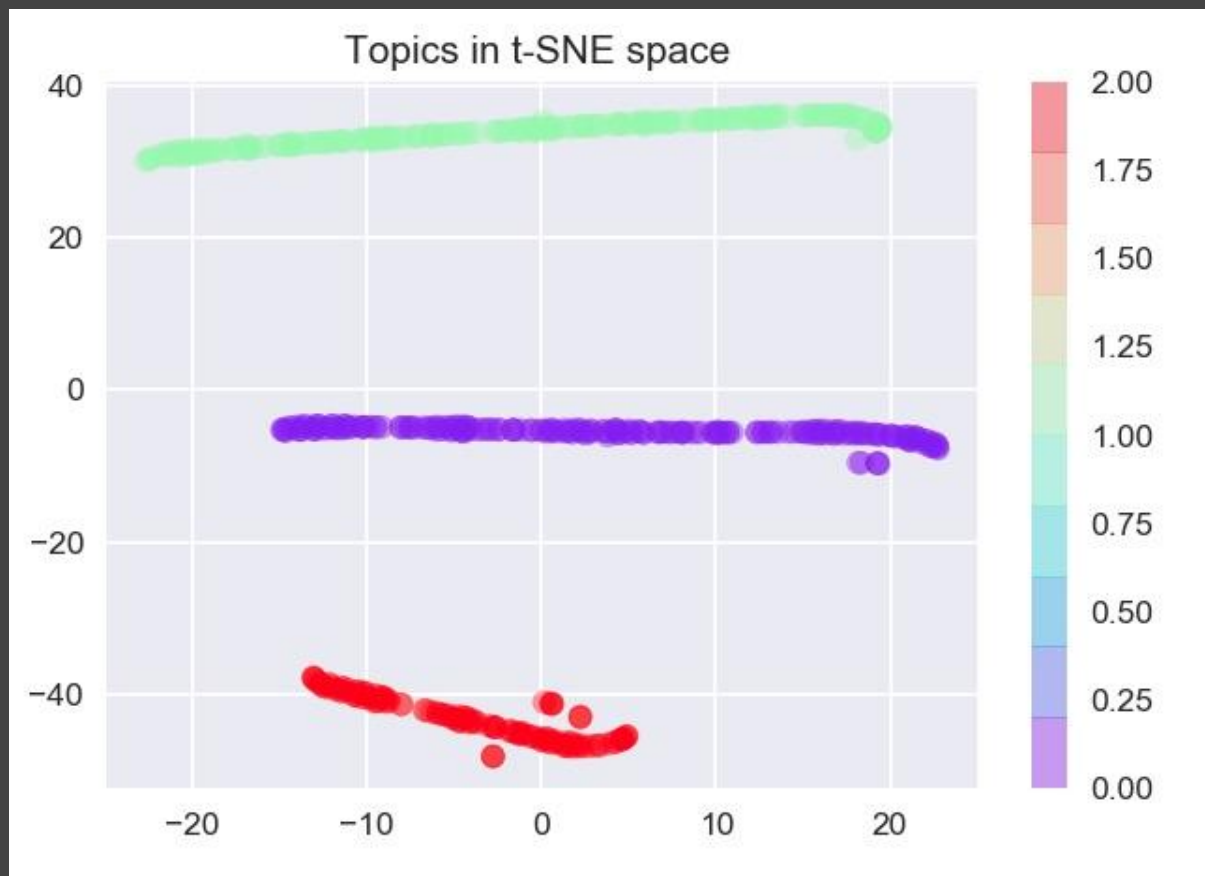
- Countvectorizer

- LDA

- Three topics

Clustering

- KMeans = 3 clusters

- TSNE confirms

Topics in KMeans clusters

Topics in t-SNE space

Interpreting the Results

- Topics correlate with sources

- Formula to measure bias based on data

Source	Right	Center	Left
Breitbart	79	48	53
Huffington Post	33	56	62
Reuters	89	120	10

Bias Scores

Source	Right	Center	Left
Breitbart	.44	.19	.36
Huffington Post	.28	.11	.62
Reuters	.33	.66	.01

Next Steps

- More data with diversified sources

- Develop more distinct topics

- Find loaded words and phrases