

The background features a gradient from dark blue on the left to bright red on the right. Numerous five-pointed stars are scattered across the field, with colors ranging from light blue to bright red. A network of thin white lines connects various points, creating a web-like structure that is more prominent on the right side.

THE DEPOLARIZATION PROJECT:

**USING DATA SCIENCE TO DISSECT AND DESEGREGATE
CONTEMPORARY POLITICAL DISCOURSE**

YOU LIKELY HAVE FELT THE IMPACTS OF POLITICAL DIVISIVENESS

As the stressors being placed on the world increase, so too do the tensions democracies across the world are feeling. Add to this the echo chambers created by social media, and it becomes easy to forget that there likely isn't all that much that separates us.



WHAT IF WE COULD USE DATA SCIENCE TO CREATE A LESS-DIVIDED POLITICAL CLIMATE?

It might seem hard to imagine today, but by better understanding what is driving the dis-unity, and along what lines we agree, it might just be possible.

As two data scientists deeply concerned about political polarization, we each decided to commit our capstone projects to advancing the dialogue, to do our part to quiet the storm, and to do what we could to use our skills to give back.

POLITICS, POLARIZATION, AND PRAGMATICS

Christine Egan



PROJECT GOALS

1. Investigate how language influences political discourse by building a model to predict political affiliation based on a sample of text.
2. Analyze the results and features of this model to make generalizations about the relationship between language and politics.

HYPOTHESES

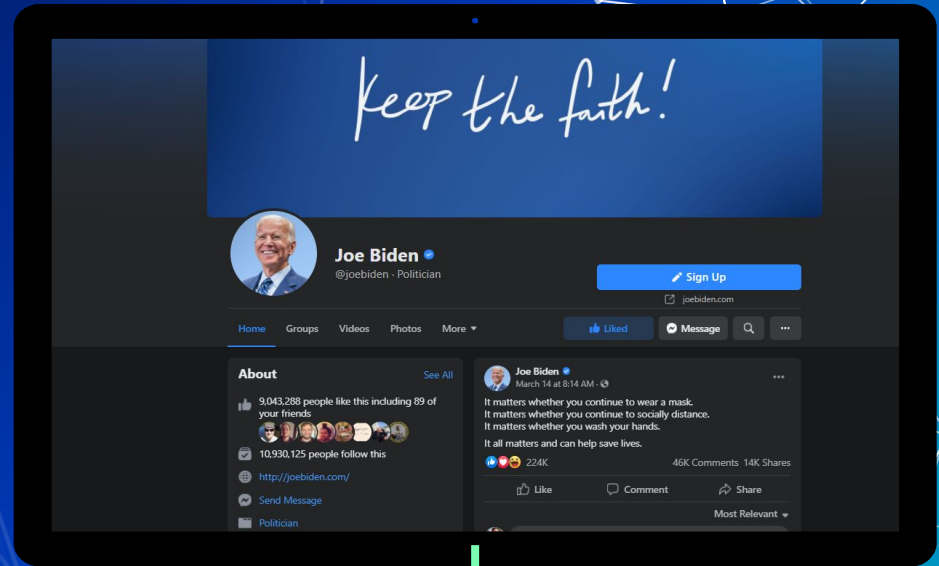
1. Language used by liberals and conservatives in political ads will be distinct.

We can test that by building a model using the language of political ads on Facebook, to see if it can distinguish between an ad with a conservative message versus a liberal message.

2. There will be some evidence of language that fit into the Strict Father/Nurturant Parent frames of Moral Politics Theory

METHODOLOGY

- Collect Facebook political ad data from ProPublica
- Label the data according to the political affiliation (liberal or conservative) that paid for the ad
- Process the ad message data
- Run a variety of models to achieve 88% accuracy in predicting if an ad was paid for by a liberal or conservative source
- Analyzed and made generalizations about the language data

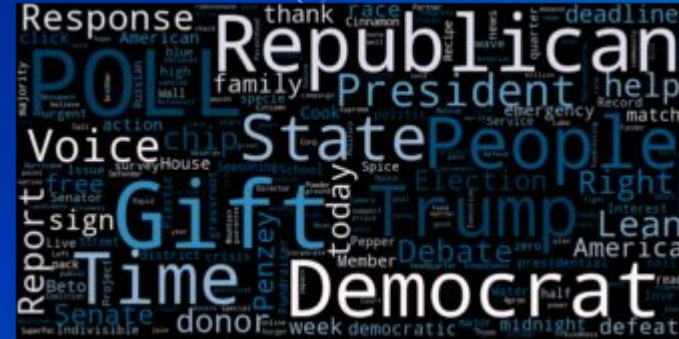


		clf	size	f1	precision	recall	accuracy	roc_auc
0	logistic regression	0.2	0.682903	0.658216	0.709514	0.674500	0.674915	
1	logistic regression	0.4	0.695192	0.651939	0.744593	0.683000	0.684736	
2	logistic regression	0.6	0.685001	0.639422	0.737577	0.667833	0.669246	
3	gaussian nb	0.2	0.863957	0.929988	0.806680	0.874500	0.873696	
4	gaussian nb	0.4	0.861119	0.896379	0.828527	0.870250	0.869074	
5	gaussian nb	0.6	0.858243	0.883285	0.834581	0.865000	0.864384	
6	support vector classifier	0.2	0.752929	0.757949	0.747976	0.757500	0.757387	
7	support vector classifier	0.4	0.731884	0.735692	0.728115	0.741000	0.740637	
8	support vector classifier	0.6	0.731642	0.722536	0.740980	0.733833	0.733978	

FINDINGS

- There seems to be a relationship between the language being used and political affiliation.
- In both groups, the name of the opposite major party appeared in the top 25 more frequent words in both groups. In other words, they spend a lot of time talking about each other.
- There is evidence of Strict Father Language in conservative samples and Nurturant Parent language in liberal samples.

Liberal Most Common Words



Conservative Most Common Words





POLITICS OF EARTH

Joe Sanders

PROJECT GOALS

1. Use unsupervised learning model to determine the political “clusters” that exist in the world today.
2. Create a supervised learning model that will predict the probability that participants are in a given cluster.
3. Create a test where participants will be able to get real-time results on which political cluster they belong to, and based on which beliefs.
4. Use the data sourced from the website test to further enrich the unsupervised learning model, making it more accurate over time.

HYPOTHESES

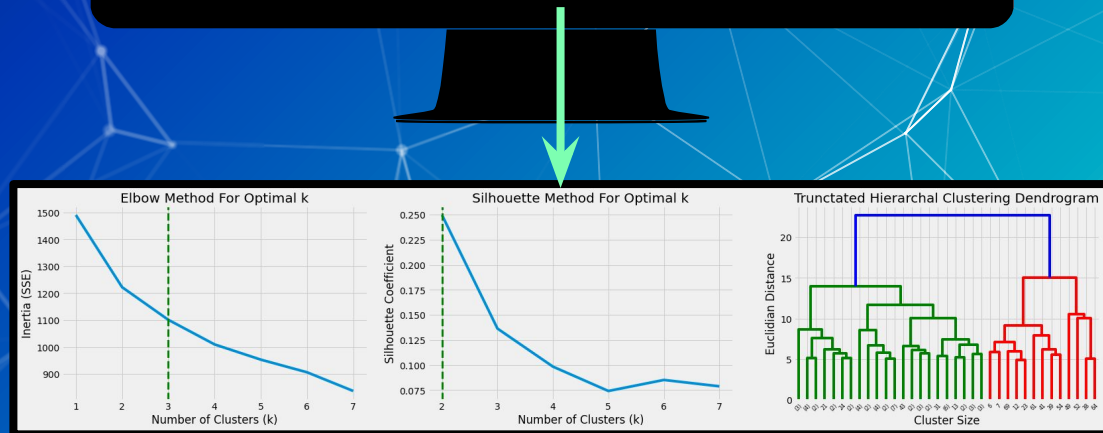
1. We will be able to clearly identify some number of political “clusters” that could be the basis of political coalitions.
2. Participants in each nation will be evenly distributed across those coalitions.
3. For any cluster, there will be at least one single-issue-based alliance with any other cluster.
4. Specific to the United States, the proposed coalitions will not naturally align with the current two-party system.

METHODOLOGY

- 64 questions spanning 11 political dimensions (Domestic Policy, Foreign Policy, Race/Gender/Sexual Orientation, etc.)
- A free-form text box for each political dimension for eventual language analysis
- Ran cluster analysis on each political dimension, and on four “decision-making” dimensions based on four dichotomies found in psychology & ethics
- Used a stepwise selection process so that only statistically relevant questions were considered (See Appendix A)

DOMESTIC POLICY: For each of the statements below, please indicate how strongly you agree or disagree.

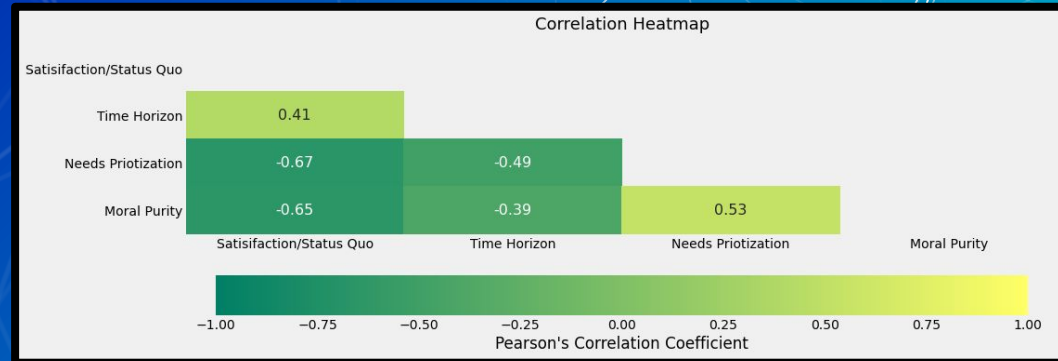
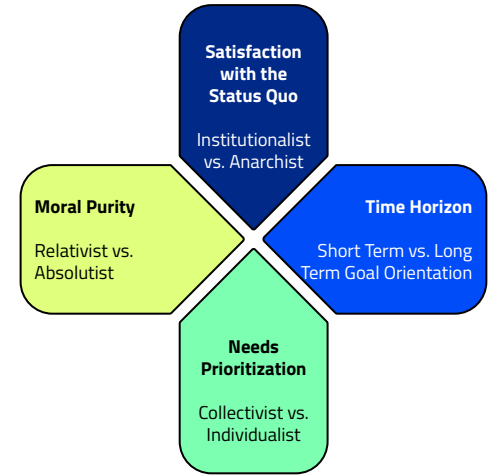
	Strongly Disagree	Disagree	Slightly Disagree	Unsure/No Opinion	Slightly Agree	Agree	Strongly Agree
There is now a worrying blurred line between information and entertainment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Taxpayers should not be expected to							



FINDINGS

- Psychology and ethics-based decision-making dimensions were more informative than politically-oriented dimensions
- Satisfaction with the Status Quo had strong correlation with Needs Prioritization ($p = -.67$) and Moral Purity ($p = -.65$).
 - Institutionalists are more likely to be Individualists and Absolutists
 - Anarchists are more likely to be Collectivists and Relativists.
- Political dimensions were distributed across decision-making dimensions

The Meeks-Sanders Political Motivations Compass



INSTITUTIONALIST VS. ANARCHIST

INSTITUTIONALIST

- "Our current process improves outputs."
- "Some people or ideas are more credible than others."
- "We have an agreed-upon set of rules we should follow."



ANARCHIST

- "The elite are corrupt/hypocritical."
- "The current system is designed to wear us down and trick us into submission."
- "No one has a monopoly on the truth."

SHORT TERM VS. LONG TERM GOAL ORIENTATION

SHORT TERM

- "We can't deal with tomorrow's challenges until we've dealt with today's challenges."



LONG TERM

- "Now is the best time to deal with tomorrow's challenges."

COLLECTIVIST VS. INDIVIDUALIST

COLLECTIVIST

- "We can further ourselves as individuals by working together."
- "Citizenship is a social contract. My country/ community should lift me up and help me succeed."

C

N

INDIVIDUALIST

- "For one person to win, another person has to lose."
- "I believe in personal responsibility."

RELATIVIST VS. ABSOLUTIST

RELATIVIST

- "What works for me may not work for someone else."
- "Tolerance is a sacred guiding principle of community building."



R

B

ABSOLUTIST

- "There is a right way and a wrong way in all things."
- "The ends justify the means."

JOINT FINDINGS: POLITICAL MOTIVATIONS & MORAL POLITICS THEORY



Strict Father

- Things are fine as they are; trust authority
- You either win or you lose
- There is a right way and wrong way in all things



Nurturant Parent

- Be weary of authority, and own your own truth
- We all win when we work together
- There are shades of grey; we must be understanding

OPPORTUNITIES FOR FUTURE RESEARCH

(Christine)

EXTEND MODEL TO CITIZEN DISCOURSE

There is a considerable amount of data harvested from Reddit that could be the basis of exploration. It would then be possible to extend the model to distinguish between Elite-to-Citizen and Citizen discourse.

DEEPER ANALYSIS OF FRAMED RHETORIC

Use Name Entity Recognition and POS tagging to identify patterns in nouns and verbs, and how that intersects with lexical items and their associated frames

RELATIONSHIP BETWEEN DEMOGRAPHICS AND CONTENT

Specifically, using the available data to see if how messages differ amongst genders, races, age groups, etc.

OPPORTUNITIES FOR FUTURE RESEARCH (Joe)

RE-EVALUATE & ITERATE WITH MORE RESPONDENTS

It was not possible to make sound statistical conclusions for each hypothesis with the amount of data gathered thus far.

DEEPER ANALYSIS OF ALL DIMENSIONS

With the large amount of quantitative and text data gathered from each respondent, further research should be conducted to see if there are meaningful connections between political and decision-making dimensions.

LANGUAGE ANALYSIS OF FREE-FORM COMMENTS

There is considerable text data that could be used in an analysis of the rhetoric used by each decision-making dimension.

THANK YOU

Let's continue this conversation...

Find Christine



Web: christine-egan.xyz

Github: [christineegan42](https://github.com/christineegan42)

LinkedIn: [/in/christine-egan-107935178](https://in.linkedin.com/in/christine-egan-107935178)

Find Joe



Github: [joesanders1010](https://github.com/joesanders1010)

Facebook: [joesanders1010](https://facebook.com/joesanders1010)

LinkedIn: [/in/joesanders](https://in.linkedin.com/in/joesanders)

Medium: [@joesanders1010](https://medium.com/@joesanders1010)

Twitter: [@joesanders1010](https://twitter.com/joesanders1010)

APPENDIX A: WHICH OPINIONS WERE STATISTICALLY RELEVANT?

■ Institutional vs. Anarchist

- "I support single-payer, universal health-care."
- "My religious values should be spread as much as possible."
- "Education is an individual's investment in themselves, therefore they should bare the cost, not the taxpayer."

■ Short Term vs. Long Term Goal Orientation

- "Stricter environmental laws and regulations cost too many jobs and hurt the economy."
- "Did you vote in the last national election?"
- "Climate change is currently one of the greatest threats to our way of life."
- "The sacrifice of some civil liberties is necessary to protect us from acts of terrorism."
- "Charity is better than social security programs as a means of helping the genuinely disadvantaged."
- "It is important that we maintain the traditions of our past."

APPENDIX A: WHICH OPINIONS WERE STATISTICALLY RELEVANT? (Continued)

■ Collectivist vs. Individualist

- "Our country has made the changes needed to give minorities equal rights."
- "Those who are able to work, and refuse the opportunity, should not expect society's support."
- "Physician-assisted suicide should be legal."
- "It is important we think in the long-term, beyond our life spans."

■ Relativist vs. Absolutist

- "The government should do more to help needy citizens, even if it means going deeper into debt."
- "My religious values should be spread as much as possible."
- "Abortion, when the woman's life is not threateed, should always be illegal."