METRIX.NEWS

Analytical research guide to the 2020 presidential election

Contents

Abstract

It is 2020, presidential election year. America is facing critical environmental, social, and governance issues that must be addressed in the next presidential term. Because Americans have increasingly lost their trust in media to report transparently, progress toward these issues has been slow.

Through primary and secondary research, we highlight a few causes that have led Americans to distrust our media including affective polarization, social media, illusion of explanatory depth, and fake / misleading news.

Metrix.News is an analytical research guide to the 2020 presidential election. With this product we aimed to create a tool that addresses the causes of distrust by offering diverse perspectives, displaying a familiar interface, refusing to collect personal user data, allowing users to challenge their knowledge gaps, and centralizing credible information.

The concept behind Metrix.News was formalized from a multi-disciplinary study consisting of dozens of expert interviews and hundreds of student interactions. Our team used design thinking methods to create the user interface and rigorous user testing to refine it.

Metrix.News is developed as a free web application that can be accessed through any browser at https://metrix.news. We implemented several analytical tools that review the language contained in political news stories. Our primary processes analyze key words, sentiment, emotions, and political bias. We use natural language processing and machine learning to achieve these analyses.

All analytic processes run successfully and the results show promise. We aggregated 100,000 news stories from a variety of political biases across 31 highly debated topics. The key word analyzer returned 1,150 important words to provide context across all topics. The sentiment and emotion output is skewed but successfully highlights extreme language.

We trained a gradient boosting classifier to successfully predict political bias within individual news stories. Our team reduced its dimensionality and optimized its parameters before testing its performance on a validation data set. It achieved 70% accuracy.

Our team believes all output performs fairly well and can provide a social benefit by providing an intuitive platform for anyone seeking to educate themselves before voting in the 2020 presidential election.

We will monitor its popularity from it’s February launch date to the presidential election in November. Thank you to everyone who supported us during any stage of this project.

Issues

It is 2020, presidential election year. America is facing critical environmental, social, and governance issues that must be addressed in the next presidential term. The majority of Americans (59%) do not trust media sources to report “fully, accurately, and fairly” according to Gallup Polls (Gallup). There are a number of factors contributing to this level of distrust.

Polarization

America seems to be more divided than ever as politicians from both the Democrat and Republican parties accuse the “other side” of hypocrisy, hatred, ignorance, and stubbornness. Recently, ordinary Americans have increasingly engaged in this divisive behavior as they express a dislike and distrust for the other party. Our resentment and hostility toward the “other side” is a phenomenon called affective polarization, which has been promoted by popular media figures who make blunt, blanket statements to invoke an emotional response.

Affective polarization should not be confused with ideological polarization which deals with the disagreement in political policy. Ideological polarization can be seen as a strength in America as we are all entitled to our own opinions which encourages diversity of thought. The two are explained in greater detail by political scientist Shanto Iyengar and colleagues (Iyengar).

Affective polarization makes it feel like we must fully commit ourselves to one party or the other, but we as individuals are not binary. The rise in political information on social media has made this feeling more prevalent.

Social Media

Social media platforms use advanced analytics to target specific information into our newsfeeds based on our prior usage. Iyengar and colleagues mention the likelihood of partisan media playing a role in the heightened animosity across party lines by stating, “exposure to partisan news makes those with extreme attitudes even more extreme” (Iyengar page 12). Social media algorithms promote confirmation bias. The ability to block out any information that does not align with one’s beliefs creates an echo chamber effect where extremism can fester.

In the 2016 presidential election, the Cambridge Analytica scam maliciously gathered and analyzed 87 million Facebook users’ data to sell psychological profiles of American voters to political campaigns in order to sway votes (Cambridge). Collecting and analyzing political user data to affect the outcome of an election is a direct threat to democracy.

Pew Research reported only 22% of all Americans are on Twitter, yet it is the hottest place for political updates (10 facts). All politicians from the president down to local officials use the platform to reach their constituents’ attention. Comments almost immediately turn to heated arguments where the goal is to slam, attack, and insult the “other side” with reactionary quips while leaving out any logical reasoning, compromise, or understanding.

Illusion of Explanatory Depth

Unless explicitly tested, most people overestimate their understanding of how political policy actually works. This phenomenon is called illusion of explanatory depth (IOED).
We all believe ourselves to be more intelligent regarding complex concepts than we actually are. IOED can be tested by asking individuals why they support a policy and then, more importantly, how that policy will bring the change they want. Confronting our gaps in knowledge has been found to make political attitudes less extreme and make people more open to contradictory information (Political Extremism).

Currently, people are prone to scroll through headlines and base their views off very limited information – especially young Americans. It is simpler to read headlines and assume they already understand the details.

Fake and Misleading Information

Most people are familiar with the concept of “fake news,” but many still fall victim to malicious content and struggle to classify credible from misleading. FactCheck.org suggests how to spot fake news and steer clear of harmful information (Kiely). We will continue to highlight a few.

“Check your biases.” Personally disagreeing with an author’s stance does not make it fake news. It is important to understand that writers take stances on issues in op-ed articles, but their opinions are still derived from real, credible information. To contrast, fake news exists to deceive and invoke quick, strong responses.

“Read beyond the headline.” Doing so can reveal common rhetorical devices and overly emotional language. Authors who use “whataboutism” are often seeking to mislead their readers. This rhetorical device was made popular by the Soviet Union during the Cold War by Joseph Stalin and is used to deflect attention away from a personal wrongdoing (What About 'Whataboutism'?). Fear and anger based tactics are common in political stories as authors attempt to rile up their bases. These tactics contribute to affective polarization as overly emotional language triggers quick responses in the audience to condition their beliefs (Marcus, G.).

“Consider the source.” Finally, it has become more difficult to recognize malicious sources since the digitalization of media. Without knowing what exactly to look for, a fake news site can look strikingly similar to real ones.

Distrust in Political Information

As mentioned in the opening paragraph, Americans do not trust our media outlets to present information “fully, accurately, and fairly.” We have only scratched the service of the biggest problems causing this distrust, and we encourage you to review our references to learn more. It is too difficult to navigate past these issues and actually learn the details necessary to form our own opinions in our busy, daily lives. We asked ourselves, is it possible to create something that addresses all these issues to help people educate themselves before the 2020 presidential election?

Hypothesis

America’s distrust in media can be addressed with a platform that...

- offers diverse perspectives
- displays a familiar interface
- refuses to collect user data
- allows users to challenge their knowledge
- centralizes credible information and data

Rationale

Metrix.News is an analytical research guide for the 2020 presidential election. We are used to organizations analyzing our personal user data, but we decided to flip the script and analyze the information they publish.

Our analytic process pulls in daily, political news stories across 31 of the most highly debated issues and each 2020 presidential candidate. All stories are sourced from a diverse set of publishers. Each story is analyzed individually, meaning publisher and author names have no influence in our classifications.

Our analytic system then breaks down each article to perform four primary analyses:

• Word Importance – A TF-IDF algorithm considers each word in each story and then all words in all stories to determine which words and phrases are most important

• Sentiment Analysis – A popular open source sentiment classifier then assigns each story to one of the following classes to assess the author’s tone: Positive, Neutral, or Negative

• Emotional Analysis – An emotional lexicon is then compared to each story to search for language indicating any of the following emotions: Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger, or Anticipation

• Political Bias Predictor – Finally, all outputs are aggregated to train an algorithm to classify the author’s political bias to one of the following classes: Left, Center-Left, Center, Center-Right, or Right.

All of this information is displayed on our intuitive user platform for readers to consider as they study the most prevalent issues facing America in 2020. Metrix.News has a similar feel to popular social media and streaming services which is appealing to a wide audience.

None of the processes or systems behind Metrix.News collect or store any personal user data. Users can safely learn without the worry of how their data is being stored or analyzed.

We also centralized links to external research organizations – government and independent – that users can review to challenge their knowledge gaps.

The chance of finding fake news is greatly reduced as we do not permit user-generated content and only pull from licensed news publishers.

Methods

The concept behind Metrix.News began in an undergraduate course at Indiana University Bloomington and was just that – a concept. Due to overwhelming support and mentorship, the concept gradually evolved into a product. Below, we outline the process behind this evolution before diving into exactly how it all works.

Primary Research

We began by conducting as much primary research as possible. In our eyes, there was clearly a problem in both the distribution and consumption habits of political media, but we needed a greater understanding of its causes. Our research was conducted using expert interviews, student interviews and surveys, and independent experimentation.

We conducted a multi-disciplinary study, by interviewing industry experts from a variety of backgrounds. Each one provided different insights to the issues facing political media and offered their own potential solutions. The interviewees came from backgrounds in journalism, psychology, political science, politics, technology, law, and business. They hold careers as university professors, journalists, state congressmembers, technical advisors, legal advisors, and venture capitalists. From these interviews, we began to notice trends in each expert's perspective and a solution began to paint itself. We grew an understanding of the issues themselves, how they impact readers, what solutions are possible, and how those solutions can be established.

We realized that we needed to reduce the scope of our study by targeting a specific demographic – at least initially. High school and college students are the most impressionable and open to contrary ideas as their political opinions are just forming. We found they are eager to engage in politics as a whole. Our team received feedback from nearly 300 students of varying ages, backgrounds, races, and ethnicities through personal interviews, surveys, roundtables, and open discussions. Again, we noticed trends in their perspectives that confirmed the plausibility of our solution. We grew and understanding of how they currently consume political media, what issues they are most passionate about, and how they engage in the political environment.

Design Thinking

We deployed design thinking methodologies which is an iterative process that seeks to understand basic user needs, challenge assumptions, and create highly intuitive products (What is Design Thinking).

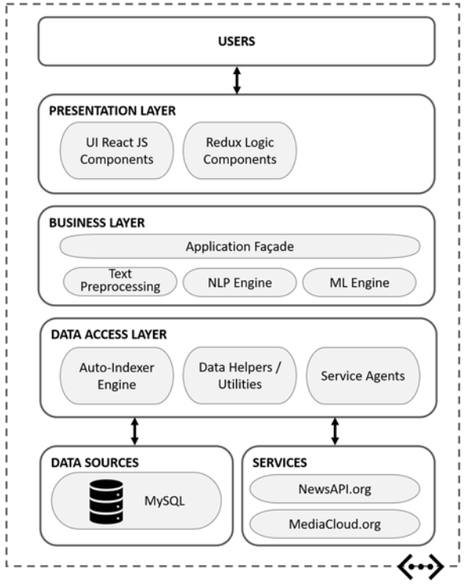
Our process began with paper and colored pencils. We provided students with a very basic template and asked them to draw their ideal news platform. The template gradually became more complex as we turned reoccurring suggestions into permanent design features. Eventually, these were made into digital proofs and then a minimally viable product. We conducted rigorous user tests on a minimally viable product and a Beta version before finalizing its current form.

Development

After establishing, designing, and proving the concept, we began investing in its development. We aimed to minimize infrastructure costs allowing us to offer the platform as a free, open source tool and maximize its potential social benefit. No processes, systems, or tools in our entire infrastructure gather or store personal user data.

Tiers & Layers

Metrix.News operates on a hybrid, 3-tier architecture. The data and business logic tier resides on a physical dual Xeon processor, 256gb RAM, 128gb hard drive machine. The application tier is hosted on a Digital Ocean virtual machine. All components are written in a model-view-controller framework. It is deployed as a web application. Consider the following chart:



Users - Access https://metrix.news through any browser on any device.

Presentation Layer - UI Components present all information to users with React JS libraries. Logic Components pass data to UI Components using Redux.

Business Layer - Application Façade allows the user interface to easily communicate with other functions.

Text Preprocessing models tokenize each news story by word, removes all stop words, and lemmatizes each remaining word.

NLP Engine performs a TF-IDF calculation, a rule-based emotional analysis, and a sentiment classification.

ML Engine considers all NLP output features along with five political bias labels on 100,000 news stories derived from MediaCloud. It trains a gradient boosting classifier using cross validation, recursive feature elimination, and hyperparameter optimization methods. During testing, the model calculates an F1_micro performance measure and plots several other performance indicators.

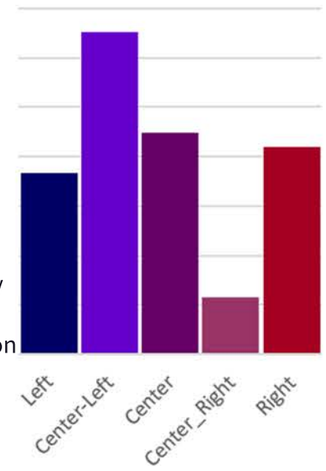
Data Access Layer - Auto-Indexer Engine, Data Helpers / Utilities, and Service Agents perform ETL processes between layers. They distribute data and information across all processes.

Data Sources and Services - MySQL relational database stores all news story data, analytic data, and routing data. MediaCloud provided 100,000 news stories and respective political bias labels to train the ML model. NewsAPI provides daily, production-ready articles to present to users.

Results

Corpus Details

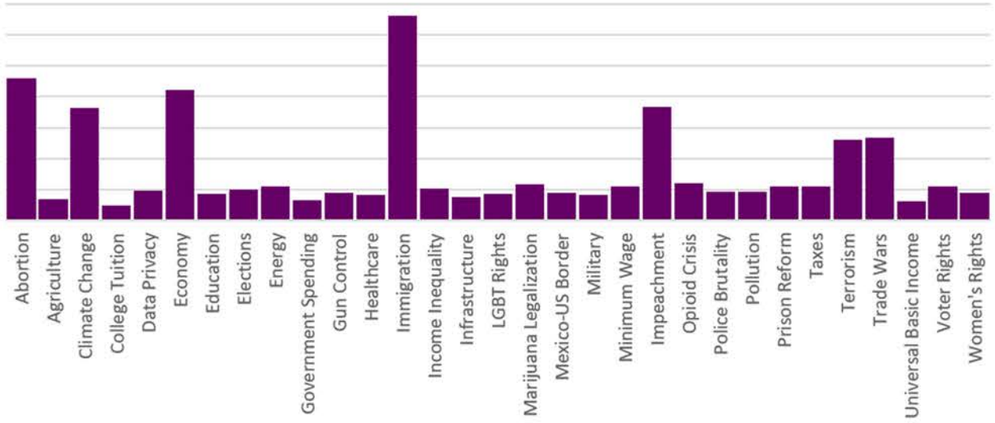
Using MediaCloud tools, we gathered 100,000 news stories from five political bias classes and 31 broad topics. The histogram to the right illustrates the distribution of political biases. The one below shows the topic distribution.



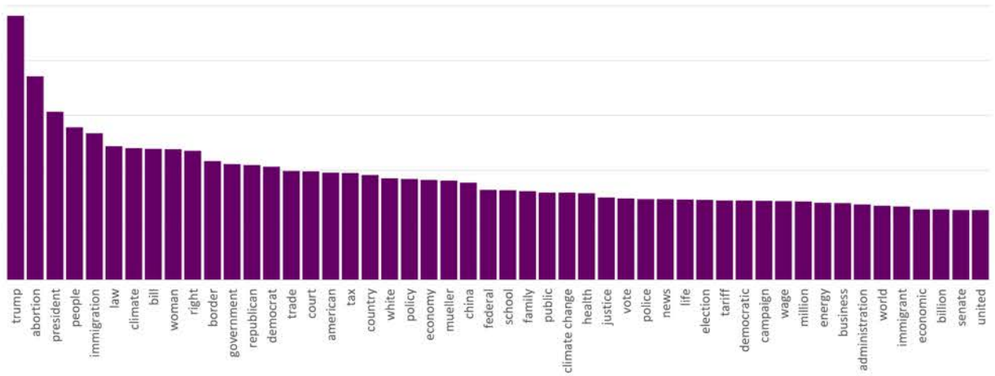
TF-IDF Output

From the text of each story, the TF-IDF model extracted 1,150 key words and phrases. Fifty of the most important words from all documents are plotted below. The most important is displayed on the left and the scores gradually reduce moving right.

Topic Distribution



TF-IDF Top 50 Word Importance Scores

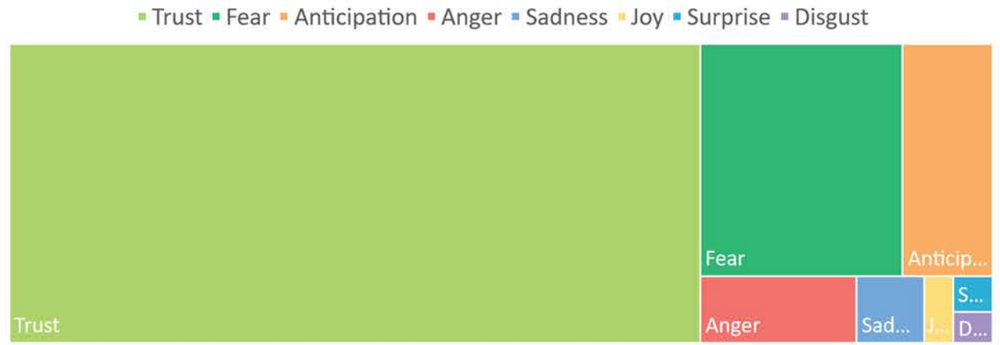


Sentiment Output

The distribution of sentiment classes among the 100,000 stories in our data set is shown in the map below. It is derived from a composite score which is often neutralized due to the structure of journalistic pieces.

Emotional Output

Due to the rules of the emotional analysis, Trust is returned as the leading emotion in the majority of stories. Fear and Anger are highlighted as well.



Political Bias Predictor

All TF-IDF, sentiment, and emotion scores are pulled together and passed to train the political bias gradient boosting model. We ran a cross validated recursive feature elimination function and hyperparameter optimization function to improve its performance. The final chart displays a confusion matrix depicting the model's predictions on the validation data set.

Recursive Feature Elimination

1,150 → **800**

Initial Features — Final Features

Hyperparameter Optimization

65% → **70%**

Initial Accuracy — Final Accuracy

	Left	Center-Left	Center	Center-Right	Right
Left	**2365**	309	1362	11	217
Center-Left	235	**2902**	346	5	148
Center	269	225	**5841**	14	356
Center-Right	181	86	897	**475**	65
Right	310	336	853	13	**2593**

Discussion

Interpretation of Results

When creating the initial corpus of news stories, some political trends emerged. There are fewer conservative news outlets than liberal, and “Center-Right” media is the most difficult to define. Also, a few major topics in 2019 skewed the distribution of topics such as immigration, abortion, climate change, and impeachment. However, media often reports in cycles, so it should be beneficial for the gradient boosting model to learn from imbalanced data to help predict political bias from stories in 2020.

The TF-IDF model extracted many important key words. We feel they are very relevant and provide some context to the political bias model.

The sentiment model classifies most stories as neutral. Journalists generally provide multiple perspectives in any given story which balances its overall tone. An author must use a very positive or negative tone to return those respective classes.

In the emotion analysis, Trust contains more words in the lexicon than any other class. However, we are pleased with it’s ability to recognize angry and fearful language as we view those as the most harmful in political media. Highlighting fear and anger based tactics can be very beneficial for users.

When combining all scores, we believe it could be a good indication of political bias. However, we understand it is naïve compared to a reader’s ability to understand bias. Mistakes will be made, but we hope it encourages users to seek a balance in information at a minimum.

User Acceptance

Currently, we do not have enough information to discuss how well our product is accepted by users. We have received positive feedback in all design thinking and user testing activities - especially in the final round. Google Analytics will track the popularity of Metrix.News from it’s launch date until the presidential election in November. This section will be updated once we have the results.

Moving Forward

Our team is financing the product with personal resources. However with an additional $500 per month, we could pull from more news sources and refresh the platform with all new content on a daily basis instead of its current weekly basis. If you would like to contribute, please contact us at info@metrix.news.

Metrix.News can provide a social benefit for all those seeking to educate themselves before the 2020 presidential election. We the people must unite to solve the critical issues facing America and the world. The first step to creating any solution is to research and understand the people affected most. We hope Metrix.News can provide the information necessary to do just that.

Metrix.News Founders

AJ Raymond
- Researcher
- Frontend Engineer
- Data Scientist

Regan Stevenson
- Mentor & Advisor

Marcus Cymerman
- Backend Engineer
- Data Scientist

Simon Mantlo
- Researcher
- Data Engineer

Acknowledgements

We sincerely thank all organizations, services, and people who contributed to this project at any stage in its development. Metrix.News could not have been created without your support. Our team is especially appreciative of the folowing organizations:

Johnson Center for Entrepreneurship and Innovation
Indiana University Kelley School of Business

Shoemaker Innovation Center
Indiana University School of Informatics

MediaCloud.org
MIT Center for Civic Media & Berkman Klein Center for Internet and Society

NewsAPI.org
HTTP REST API for searching and retrieving live articles from all over the web

NRC Emotion Lexicon
Saif M. Mohammad - National Research Council Canada

JoyPixels
Full service emoji provider

References

10 facts about Americans and Twitter. (2019). Pew Research Center. Retrieved 11 February 2020, from https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/

Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. (2020). Nytimes.com. Retrieved 11 February 2020, from https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Gallup, I. (2019). Americans' Trust in Mass Media Edges Down to 41%. Gallup.com. Retrieved 11 February 2020, from https://news.gallup.com/poll/267047/americans-trust-mass-media-edges-down.aspx

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., Westwood, S. (2020). Dartmouth.edu. Retrieved 11 February 2020, from https://www.dartmouth.edu/~seanjwestwood/papers/ARPS.pdf

Kiely, E., & Robertson, L. (2016). How to Spot Fake News. FactCheck.org. Retrieved 11 February 2020, from https://www.factcheck.org/2016/11/how-to-spot-fake-news/

Marcus, G., Pérez, F., Quizar, J., Editors, I., & Labuski, C. (2020). How Fear and Anger Impact Democracy. Items. Retrieved 11 February 2020, from https://items.ssrc.org/democracy-papers/how-fear-and-anger-impact-democracy/

Political Extremism is Supported by an Illusion of Understanding. Fernbach, P., Rogers, T., Fox, C., Sloman, S. (2020). Meteo.mcgill.ca. Retrieved 11 February 2020, from http://www.meteo.mcgill.ca/~huardda/articles/fernbach13.pdf

What About 'Whataboutism'?. (2020). Merriam-webster.com. Retrieved 11 February 2020, from https://www.merriam-webster.com/words-at-play/whataboutism-origin-meaning

What is Design Thinking and Why Is It So Popular?. (2020). The Interaction Design Foundation. Retrieved 11 February 2020, from https://www.interaction-design.org/literature/article/what-is-design-thinking-and-why-is-it-so-popular