

Wahl-O-Mat®

Bundestagswahl 2021

1/38 Voter's interests and Wahl-O-Mat Questions

Does the Wahl-O-Mat influence election outcomes?

stimme zu

neutral

stimme nicht zu

[These überspringen →](#)



A topic analysis of Wahl-O-Mat questions

By Anna Engler, Tristan Engeler and Miel Giménez Weber

Group project in Python for Economics and Social Sciences

Seminar at Sciences Po Paris by Moritz Pfeifer

10.12.2021

1) Introduction

The Wahl-O-Mat is an online voting advice application run by the German “Bundeszentrale für Politische Bildung” (the German agency for political education) and was first created for national parliamentary elections in 2002. The official vote compass presents to its users a set of 38 or 40 questions of political relevance that these can agree, disagree with, be neutral about or skip and even assign a heightened importance to. The tool then continues to rank the parties most in line with the users’ answers on upcoming national or federal parliamentary elections.

The questions are created before each election by a committee of 20 young voters, political scientists, statisticians and are overseen by the “Bundeszentrale für Politische Bildung”. Political parties of all sizes are then able to answer the questions, which will, in the end, lead to a voting recommendation. Several alternatives to the Wahl-O-Mat exist, such as a multilingual version or one solely dealing with parties’ positions on climate change but are in turn not financed and legitimized by the state authority.

The official Wahl-O-Mat usually gets published between two to four weeks in advance of elections and has so far been used more than 100 million times, including 21 million times before national parliamentary elections in 2021. More than 30% of its users are under the age of 30, and over one third have at least completed a baccalaureate. About 15% of its users describe themselves as not being interested in politics.¹

There is a non conclusive academic debate about the question whether voting advice applications (VAA) have a significant influence on voting turnout and voting choice. A field study from 2017 concludes that VAA do not significantly influence voting turnout and choice but instead only increase citizen’s knowledge about the parties’ positions.² This study has a relatively small sample size of only 1000 people compared to a population of 61,5 million that was eligible to vote in Germany at the time. However, a 2021 meta-analysis of 22 studies in nine European countries found that VAA has a significant influence on vote choice and can similarly increase voter turnout, as those who declared themselves abstentionists were about 20% more likely to cast a ballot after having completed the Wahl-O-Mat, ascribing the VAA an “activation effect”.³

Given the findings of this meta-analysis it seems crucial to analyze whether the topics curated for the questions, and on the basis of which voting advice is given, are in line with the topics that are relevant to the population in question. Assuming that the Wahl-O-Mat is able to influence user’s voting schemes, we deduce that the Wahl-O-Mat is also indirectly able to influence election results. The first startling discrepancy we noticed without writing a single line of code or analyzing statistics, was that for the 2021 national elections there was a single question about Covid in the Wahl-O-mat, although for us personally it’s one of the major concerns.

¹ Bundeszentrale für politische Bildung, “Die Nutzer Des Wahl-O-Mat,” September 27, 2021, <https://www.bpb.de/politik/wahlen/wahl-o-mat/177430/die-nutzer-des-wahl-o-mat>.

² Simon Munzert et al., “Do Online Voter Guides Empower Citizens? Evidence from a Field Experiment with Digital Trace Data,” *Public Opinion Quarterly*, Fall 2020, Volume 84 edition, <https://academic.oup.com/poq/article/84/3/675/6104133>.

³ Simon Munzert and Sebastian Ramirez-Ruiz, “Meta-Analysis of the Effects of Voting Advice Applications,” *Political Communication*, January 18, 2021, 1–16, <https://doi.org/10.1080/10584609.2020.1843572>.

2) Methodology

2.1) Presentation of models

In order to compare the relevance of the Wahl-O-Mat questions with the general interests of the German population, we compared the Wahl-O-Mat questions with the Eurobarometer surveys, a set of studies on behalf of the European Commission that shows European citizens opinion concerning major political and social issues. Therefore, we used the data provided by the “Bundeszentrale für Politische Bildung”⁴ and extracted all of the Wahl-O-Mat questions of each federal election from 2002 to 2021. Thus, we got a dataset of 209 questions. Accordingly, we used the Eurobarometer study for each corresponding year and extracted the responses of German citizens to the question “What do you think are the two most important issues facing our country at the moment?”⁵

After that, we aimed at analysing which topics were the most relevant in the Wahl-O-Mat dataset. Therefore, we used different approaches to Topic Modeling, an unsupervised machine learning method in the field of natural language processing (NLP), that allows us to find topics which represent a set of documents. Using Topic Modeling has the advantage that the results are not distorted by human biases, but the results of a neutral machine learning process.

First, we preprocessed our dataframe in order to make the data more accessible for Topic Modeling. We put the whole text in lower case and split the text in pieces of single words via tokenization. Then, we removed punctuations, special characters and stopwords, trivial words that occur so frequently that they distort NLP. We also used lemmatization, a process that converts the words of our dataset to their base by removing affixes, another process that improves NLP.

After applying vectorization, we could now try different Topic modeling methods on our preprocessed data. We started with Latent Dirichlet Allocation (LDA), the most common method of Topic Modeling. LDA calculates the distribution of words in each document (in our case each Wahl-O-Mat question) and assigns this distribution to generated topics. So each topic is composed out of a cluster of words. The number of topics and the number of words in each cluster can be manually determined.

Then, we used Non-negative Matrix Factorization (NMF), a method that creates a matrix composed of all of the documents and all of the words and shows the frequency of each word in each document. According to the word distribution, NMF also creates topics and calculates how well each document fits each topic.

⁴ Bundeszentrale für politische Bildung, “Wahl-O-Mat-Archiv: Weitere Wahlen,” May 7, 2021, <https://www.bpb.de/politik/wahlen/wahl-o-mat/45817/weitere-wahlen>.

⁵ Eurobarometer Data Service, “Most Important Issues,” accessed December 10, 2021, <https://www.gesis.org/en/eurobarometer-data-service/search-data-access/eb-trends-trend-files/list-of-trends/polit-issues-national>.

A third approach was the Gibbs Sampling Dirichlet Mixture Model (GSDMM). This modified version of the LDA algorithm is particularly suited for short texts because it makes the preliminary assumption that each document only contains one topic.⁶

Finally, we used BERT, a keyword extraction model that generates the most relevant words and phrases of a dataset.⁷ As BERT uses artificial intelligence it is used in many different fields such as sentiment analysis but it can also be used for topic analysis. Contrary to topic modeling, BERT is a supervised not an unsupervised machine learning technique. That means that the algorithm was already trained before with a big data-set. Hence it is very effective for text classification. With BERT, you can also manually determine the number of keywords which characterize the text.

2.2) Comparison of the different models

Researchers of the university of British Columbia found that NMF delivers better results than LDA because it generates more interpretable topics.⁸ Furthermore, a comparative analysis of applying LDA and NMF on ABC News Headlines also showed that NMF performs much better than LDA.⁹

We could confirm these observations in our analysis of the Wahl-O-Mat dataset. With the same amount of words per topic, the words in the clusters of the LDA model were much less coherent than in the NMF clusters and it was more often difficult to identify a topic. In addition, with LDA it was not possible to differentiate the weight of the single words in each topic.

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights
0	zentral	1.1	kindern	1.1	preis	1.1	homeoffice	1.1	news	1.1	betrieb	1.1
1	genetischer	1.1	andere	1.1	autofahren	1.1	ob	1.1	löschen	1.1	pipeline	1.1
2	dna	1.1	traditionelle	1.1	heizen	1.1	arbeiten	1.1	hingewiesen	1.1	nord	1.1
3	erfasst	1.1	mutter	1.1	beim	1.1	beschäftigten	1.1	falschinformationen	1.1	transportiert	1.1
4	straftätern	1.1	vater	1.1	festgelegte	1.1	selbst	1.1	internetseiten	1.1	ostsee	1.1

Result for LDA: all words have the same weight.

NMF provided different weights which made it easier to understand the relevance of each word in a cluster and made it possible to identify words that did not really fit into one cluster. This was especially necessary when we chose to display more than five words per cluster. Therefore, it was more difficult to interpret the LDA results.

⁶ Matyas Amrouche, "Short Text Topic Modeling" (Towards Data Science, August 2019), <https://towardsdatascience.com/short-text-topic-modeling-70e50a57c883>.

⁷ Nils Reimers, "Pretrained Models for Sentence Transformers" (SBERT.net, 2021), https://www.sbert.net/docs/pretrained_models.html.

⁸ May Young and David Johnson, "Comparison of LDA and NMF for Topic Modeling on Literary Themes," accessed December 10, 2021, https://wiki.ubc.ca/Course:CPSC522/A_Comparison_of_LDA_and_NMF_for_Topic_Modeling_on_Literary_Topics.

⁹ Ravish Chawla, "Topic Modeling with LDA and NMF on the ABC News Headlines Dataset" (Medium, July 2017), <https://medium.com/ml2vec/topic-modeling-is-an-unsupervised-learning-approach-to-clustering-documents-to-discover-topics-fdfbf30e27df>.

The poorer performance of LDA is due to the small size of our dataset. NMF is better suited for short texts because it calculates how well a document fits a certain topic instead of assuming that each document has multiple topics.

However, NMF and LDA both did not provide a valuation of the different clusters among themselves. Only the GSDMM model provided a ranking of the clusters according to their relevance for the entire dataset. The GSDMM model also gave better interpretable results with regard to the weight meaning the number of occurrences of the single words in each cluster and their coherence in terms of content.

Number of documents per topic : [(0, 16), (1, 20), (2, 18), (3, 18), (4, 24), (5, 20), (6, 25), (7, 21), (8, 22), (9, 25)]
 Most important clusters (by number of docs inside): [9 6 4 8 7 5 1 3 2 0]

Cluster 9 : [(['deutschland', 9), ('fur', 5), ('europa', 5), ('union', 4), ('mehr', 3), ('bund', 3), ('austret', 3), ('verbot', 3), ('schulpolit', 2), ('bedingungslos', 2)]

Cluster 6 : [(['gesetz', 10), ('all', 8), ('generell', 5), ('tempolimit', 5), ('autobahn', 4), ('einfuhr', 4), ('burg', 3), ('krankenkass', 3), ('versichert', 3), ('gelt', 3)]

Cluster 4 : [(['abgeschafft', 5), ('weiterhin', 5), ('konn', 4), ('unternehmen', 3), ('vollstand', 3), ('entscheidung', 2), ('solidaritatzuschlag', 2), ('fluchtling', 2), ('durf', 2), ('leistung', 2)]

The GSDMM results indicate the order of relevance of the topics and the frequency of the words.

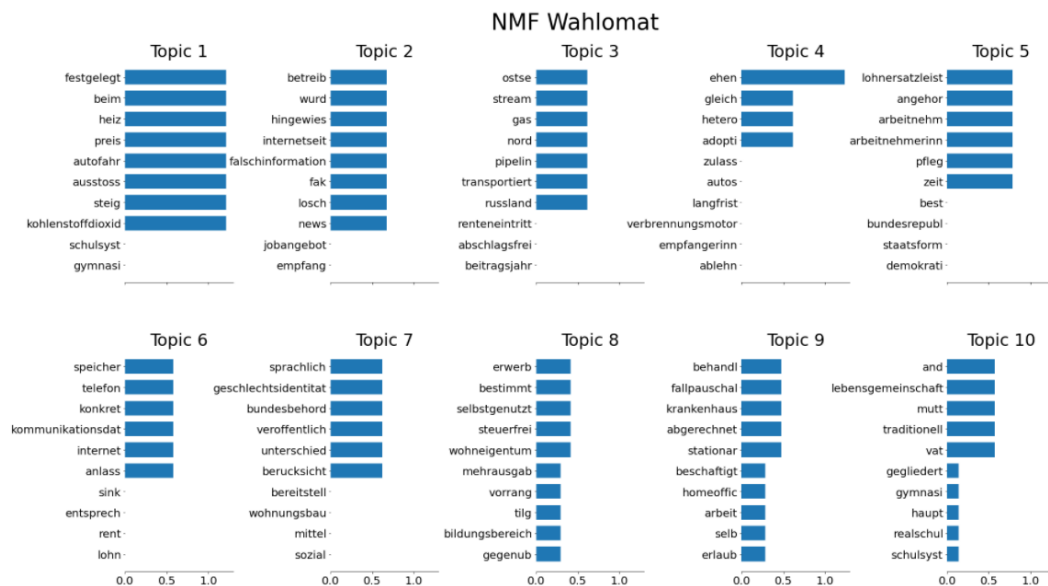
BERT gave us also more sound results as it produced a certain number of keywords which best characterize the whole dataset. As one does not see large clusters it is easier to see quickly which are the most recurrent topics. The keywords are also ranked with regards to their relative importance. However it is not possible to see exactly how many other documents (i.e. Wahl-O-Mat questions) are linked to the topic.

Finally, while due to our small dataset the common topic models LDA and NMF yield limited results, the GSDMM model and the BERT text classification transformer put forward more detailed and significant topics.

2.3) Changing the parameters

Each model we used, allows to change its parameters, like the number of topics, words and or keywords, which leads to different results.

In general, the resulting topics were more coherent with a smaller number of topics. For example, in the NMF model, it was not useful to use more than 10 topics because with a larger number of topics, more words inside the clusters had no weight at all, meaning that they were not relevant for the topic. The same is true for the number of words per cluster. The results were better interpretable with a smaller number of words (between 5 and 10 words, depending on the model).



NMF with 10 topics and 10 words per topic is not efficient because some words have no weight.

For topic classification with BERT we found the most significant results using bigrams. The disadvantage of unigrams is that some of the resulting keywords had no relevance for our analysis of the content of the Wahl-O-Mat. Furthermore, unigrams only provide relatively vague topics. The topics composed of bigrams and trigrams were more precise because they were composed out of several consecutive keywords. However, using trigrams, some resulting words were adjectives or prepositions that had no relevance for our analysis. Therefore, we opted for using bigrams.

Results for BERT for ten topics:

Using unigrams

```
[ 'autobahn',
  'tempolimit',
  'rechtsextremismus',
  'deutschland',
  '16',
  'kommunikationsinfrastruktur',
  'bildungsbereich',
  'all',
  'migrationshintergrund',
  'energieintensiv']
```

Using bigrams

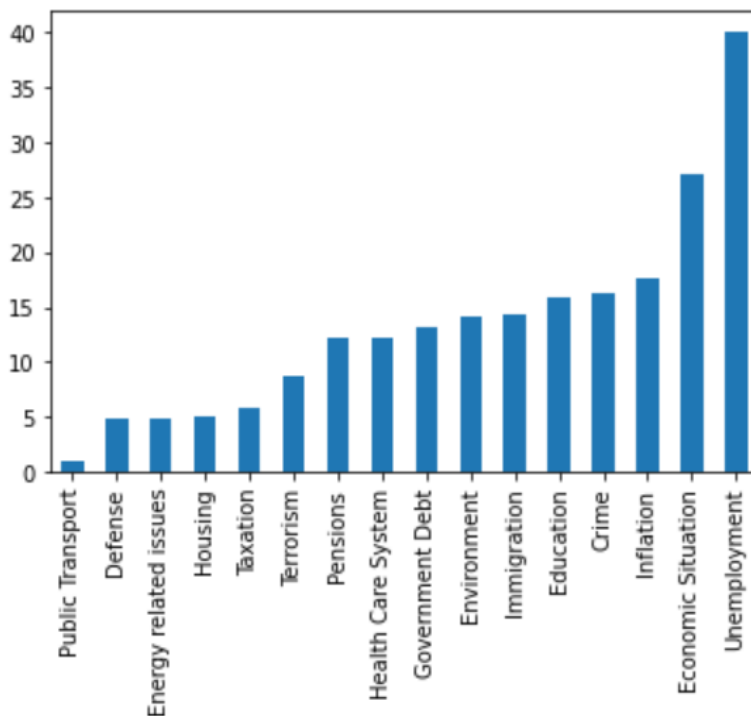
```
[ 'tempolimit autobahn',
  'deutschland verteidigungsausgab',
  'all kind',
  'ausbau kommunikationsinfrastruktur',
  'weiterhin verbot',
  'fingerabdruck personalausweis',
  'euro wahrung',
  'acht jahr',
  'geschlechtsidentitat sprachlich',
  'polizei innerhalb']
```

Using trigrams

```
['tempolimit deutsch autobahn',  
'fur all kind',  
'geg rechtsextremismus ford',  
'armst land welt',  
'speicher kommunikationsdat telefon',  
'acht jahr erhocht',  
'comput sicherheitsbehord verbot',  
'all autobahn generell',  
'unterschied geschlechtsidentitat sprachlich',  
'weiterhin zentral bestandteil']
```

3) Results

Combining the results of the Eurobarometer surveys from 2002 to 2021, we found that the most important issues for the German population regarding their country were unemployment and the economic situation. To a lesser extent, inflation, crime, education, immigration and the environment were also relatively relevant.



Results of the Eurobarometer survey 2002 – 2021

However, these issues are not the most recurring topics in the Wahl-O-Mat questions. Using the BERT approach with bigrams and 10 topics, we found that the Wahl-O-Mat was particularly focused on mobility since the most relevant bigram was “tempolimit autobahn”. We also found these words in the most important cluster of the GSDMM Model which confirms their increased relevance for the Wahl-O-Mat. In comparison with the Eurobarometer, public transport is the least important issue for the German population. Therefore, we concluded that mobility was largely overrepresented in the Wahl-O-Mat.

BERT bigrams with ten topics (keywords recurring in GSDMM are color coded in yellow)

```
['tempolimit autobahn',  
'deutschland verteidigungsausgab',  
'all kind',  
'ausbau kommunikationsinfrastruktur',  
'weiterhin verbot',  
'fingerabdruck personalausweis',  
'euro wahrung',  
'acht jahr',  
'geschlechtsidentitat sprachlich',  
'polizei innerhalb']
```

GSDMM cluster with 10 topics and 10 words per topic (keywords recurring in BERT are color coded in yellow)

```
Cluster 3 : [(['erhoht'], 6), (['verbot'], 5), (['deutschland'], 5), (['generell'], 5), (['sollt'], 4), (['tempolimit'], 4), (['autobahn'], 4), (['deutsch'], 4), (['jahr'], 3), (['gesenkt'], 3)]  
  
Cluster 8 : [(['fur'], 7), (['stark'], 6), (['bund'], 5), (['ökolog'], 4), (['landwirtschaft'], 4), (['finanziell'], 4), (['staat'], 3), (['geg'], 3), (['gefordert'], 2), (['konventionell'], 2)]  
  
Cluster 7 : [(['all'], 8), (['gesetz'], 6), (['unabhäng'], 5), (['einkomm'], 5), (['elt'], 5), (['bafög'], 3), (['gezahlt'], 3), (['burg'], 3), (['krankenkasse'], 3), (['versichert'], 3)]  
  
Cluster 5 : [(['abgeschafft'], 4), (['kind'], 4), (['erhalt'], 4), (['vollstand'], 3), (['elt'], 3), (['fur'], 3), (['staatlich'], 3), (['solidaritätszuschlag'], 2), (['betreuungsgeld'], 2), (['end'], 2)]  
  
Cluster 2 : [(['deutschland'], 11), (['deutsch'], 6), (['dürf'], 5), (['weiterhin'], 4), (['europa'], 4), (['fur'], 4), (['staatsangehör'], 3), (['union'], 3), (['neb'], 2), (['zweit'], 2)]  
  
Cluster 0 : [(['eingeführt'], 4), (['einfuhr'], 4), (['gesetz'], 4), (['flächendeck'], 4), (['z'], 3), (['b'], 3), (['volksentscheid'], 3), (['bundeseib'], 3), (['mindestlohn'], 2), (['kind'], 2)]  
  
Cluster 6 : [(['dürf'], 5), (['deutschland'], 4), (['deutschland'], 3), (['ohn'], 3), (['frau'], 2), (['vorher'], 2), (['berat'], 2), (['straffrei'], 2), (['abtreib'], 2), (['austret'], 2)]  
  
Cluster 4 : [(['offent'], 5), (['dürf'], 4), (['eingesetzt'], 4), (['fur'], 4), (['staat'], 4), (['videouberwach'], 3), (['weiterhin'], 3), (['terrorismusbekämpfung'], 2), (['bundeswehr'], 2), (['konn'], 2)]  
  
Cluster 1 : [(['fur'], 4), (['konn'], 3), (['jahr'], 3), (['unternehmen'], 2), (['entscheid'], 2), (['uber'], 2), (['neu'], 2), (['straftat'], 2), (['europa'], 2), (['gesenkt'], 2)]  
  
Cluster 9 : [(['erlaubt'], 4), (['generell'], 3), (['fur'], 3), (['unternehmen'], 3), (['weiterhin'], 3), (['kontrolliert'], 2), (['verkauf'], 2), (['cannabis'], 2), (['abgeschafft'], 2), (['gesenkt'], 1)]
```

Other strongly recurring issues of the Wahl-O-Mat were security, defense and crime related to the keywords that BERT gave in 3 out of 10 bigrams: “verteidigungsausgaben”, “polizei”, “fingerabdruck” and “personalausweis”. This topic corresponds to cluster 4 of the GSDMM Model including video surveillance, counterterrorism and the German army. In comparison with the Eurobarometer, these issues are also rather overrepresented since crime, and especially terrorism and defence are not ranked under the most important issues of the German population. The economic situation was also a relevant topic for the Wahl-O-Mat as we could observe from BERT with the 7th bigram “euro wahrung” and the GSDMM cluster 0 about the introduction of a nationwide minimum wage. This corresponds to the results of the Eurobarometer, showing the high relevance of the economic situation for German citizens.

Another topic we could find was identity politics, related in BERT to the 9th bigram “geschlechtsidentitat sprachlich” and in GSDMM to the cluster 6 concerning women, abortion and impunity. This topic is completely absent from the Eurobarometer study. The same is true for the fourth bigram “aufbau kommunikationsinfrastruktur” which refers to the theme of digitalisation.

Nevertheless, we cannot conclude that identity politics and digitalisation are of no interest at all for the German population since the Eurobarometer is biased by the topic-setting of the authors of this survey. For instance, the people interviewed got a prefabricated list of topics of which they could choose the one most relevant to them. Since identity politics and digitalisation were not among them, we cannot determine their relevance for the German population.

We could also observe that some of the most important issues for the German population are highly underrepresented in the Wahl-O-Mat. Unemployment, the most relevant issue by far, is not directly addressed in the questions. Inflation and the environment are also not directly identifiable in the results of BERT and GSDMM.

4) Interpretation of the results

We can thus conclude that the topics the German electorate cares about are neither perfectly overlapping nor proportionally equivalent to their representation in the Wahl-O-Mat theses. Some differences between the Eurobarometer surveys and the Wahl-O-Mat questions can be explained by the purpose and functioning of the Wahl-O-Mat. The Wahl-O-Mat aims to show the parties' divergent positions on different issues, therefore it preferably uses questions on which important German parties have contrasting opinions. Speed restrictions on highways, for example, could be an overrepresented question because it is a controversial one. Another explanation for the differences is that the Wahl-O-Mat tries to ask questions that are not too complicated such that everyone can understand and answer them. For instance, subjects such as inflation and unemployment are underrepresented in the Wahl-O-Mat because financial politics, like targeting interest rates or quantitative easing, are not easily accessible for all of the population.

Given that the voter's main interests (such as economic stability and unemployment) are underrepresented and controversial questions overrepresented, voters might be advised to vote a party they agree with on the polemic questions instead of those they agree with on issues as important as unemployment, that are less controversial. Thus, the Wahl-O-Mat might give misleading voting advice and alter election results. To compensate for this effect slightly, the Wahl-O-Mat already gives the possibility to weigh the questions.

Further research could be conducted to differentiate the results for each federal election. Our approach only compared all of the theses of the past 20 years to the sum of the Eurobarometer Study results in the same timeframe. It would be interesting to do the Topic Modeling or BERT approach for each election separately and to compare the results with the corresponding Eurobarometer survey. This would allow us to have a more detailed analysis of the differences between the topics set by the Wahl-O-Mat and the citizen's interests because their interests are evolving over time. Another differentiation could be made between West and East Germany. In our analysis of the Eurobarometer, we found that these populations' interests diverge for some topics.

5) Bibliography

- Amrouche, Matyas. "Short Text Topic Modeling." Towards Data Science, August 2019.
<https://towardsdatascience.com/short-text-topic-modeling-70e50a57c883>.
- Bundeszentrale für politische Bildung. "Die Nutzer Des Wahl-O-Mat," September 27, 2021.
<https://www.bpb.de/politik/wahlen/wahl-o-mat/177430/die-nutzer-des-wahl-o-mat>.
- Bundeszentrale für politische Bildung. "Wahl-O-Mat-Archiv: Weitere Wahlen," May 7, 2021.
<https://www.bpb.de/politik/wahlen/wahl-o-mat/45817/weitere-wahlen>.
- Chawla, Ravish. "Topic Modeling with LDA and NMF on the ABC News Headlines Dataset." Medium, July 2017. <https://medium.com/ml2vec/topic-modeling-is-an-unsupervised-learning-approach-to-clustering-documents-to-discover-topics-fdfbf30e27df>.
- Eurobarometer Data Service. "Most Important Issues." Accessed December 10, 2021.
<https://www.gesis.org/en/eurobarometer-data-service/search-data-access/eb-trends-trend-files/list-of-trends/polit-issues-national>.
- Munzert, Simon, Pablo Barbera, Andrew Guess, and JungHwan Yang. "Do Online Voter Guides Empower Citizens? Evidence from a Field Experiment with Digital Trace Data." *Public Opinion Quarterly*, Fall 2020, Volume 84 edition.
<https://academic.oup.com/poq/article/84/3/675/6104133>.
- Munzert, Simon, and Sebastian Ramirez-Ruiz. "Meta-Analysis of the Effects of Voting Advice Applications." *Political Communication*, January 18, 2021, 1–16.
<https://doi.org/10.1080/10584609.2020.1843572>.
- Reimers, Nils. "Pretrained Models for Sentence Transformers." SBERT.net, 2021.
https://www.sbert.net/docs/pretrained_models.html.
- Young, May, and David Johnson. "Comparison of LDA and NMF for Topic Modeling on Literary Themes." Accessed December 10, 2021.
https://wiki.ubc.ca/Course:CPSC522/A_Comparison_of_LDA_and_NMF_for_Topic_Modeling_on_Literary_Themes.