# Classifying Newspaper Bias with Cross-Domain Learning

Natural Language Processing, Cross-Domain Learning, Transformer Models, Newspaper Bias

Can we use party communication to train transformer models detecting newspaper bias? The short answer: no.

AUTHORS

Nicolai Berk

Tom Arend

PUBLISHED

May 3, 2021

## Problem

Often, political scientists want to measure ideological bias in political texts, such as newspapers. While supervised techniques provide an efficient tool to address this problem, we often lack available labels to train these supervised models. Additionally, the labelling of political slant is often not straightforward and affected by coders' biases.

We propose to use information from political texts with available labels - namely the press releases of political parties - to assess the relative placement of political texts with transformer models (Devlin et al. 2019). This cross-domain approach should enable researchers to assess the political bias in content that is costly or hard to label. Moving beyond existing approaches which infer political bias from the issuing actors (Widmer, Ash, and Galletta 2020) or specific phrases used by politicians (Gentzkow and Shapiro 2010), we apply state-of-the-art transformer models to measure the similarity of newspaper articles to the language of political parties.

## Task

To understand how information from political actors can be used to understand ideological bias in other context, we train several DistilBERT transformer models [1] on a collection of over 40,000 German party press releases issued between 2013 and 2019, collected by the SCRIPTS project [2] . These are then applied to a random subsample of 4,000 German newspaper articles from six major newspapers, published between 2013 and 2019, collected by one of the authors in a previous project [3] . To validate whether the estimates from our classifiers conform to general expectations about the ideological bias of these newspapers, we use data from a survey of newspaper readers asking them about the partisan bias of their newspaper (Roßteutscher et al. 2019).

## Proposed method

We assess the use of cross-domain learning to identify newspaper slant in two main studies [4] . A pre-trained transformers model is fine-tuned on our set of press releases to identify the authoring party.

In study 1, we use the DistilBERT model with different training and test sets and compare performance: first, we train on the full set of releases issued between 2013 and 2019 and assess the performance on a test set randomly chosen from this data [5] . Second, to exclude assess the performance on a test set randomly chosen from this data . Second, to exclude the possibility that the model only infers from references to the political parties, we 'blindfold' the classifier by censoring any references to the parties. Lastly, we split training and test set temporally, meaning the classifier needs to predict the labels of press releases published after the period from which the training data was sampled. If the model picks up general patterns about the parties' language, the performance on this set should only decrease slightly, meaning that the model should correctly predict the authoring party of the press release most of the time. Comparing performance using these different training and validation sets, we can assess whether the classifiers perform as expected.

In study 2, these models are applied to a range of newspaper articles, indicating which parties' communication an article most resembles. These estimates are compared to the expectations about newspaper slant formulated by the survey respondents. Ideally, the models show little respnsiveness to changes in the input data and closely resemble the survey estimates.

# Results

**Study 1**:

As described, we start by assessing the model's performance on a set of press releases issued in the same period as the training data. The results are shown in the table. As the reader can see, the model performs at a very high, near-perfect level for all categories. The impressive performance of this model on the press releases was rather surprising. Indeed, this performance seemed too good to be true and raised a number of concerns, most notably regarding issues of overfitting.

| | label | class | f1 | precision | recall | n |
|---|---|---|---|---|---|---|
| **3** | 3 | SPD | 0.997927 | 0.997238 | 0.998617 | 1446 |
| **4** | 4 | Linke | 0.997237 | 0.997543 | 0.996931 | 1629 |
| **5** | 5 | FDP | 0.996416 | 0.996813 | 0.996019 | 1256 |
| **0** | 0 | Greens | 0.995416 | 0.996940 | 0.993898 | 1311 |
| **1** | 1 | Union | 0.990737 | 0.984224 | 0.997336 | 1126 |
| **2** | 2 | AfD | 0.989836 | 0.998423 | 0.981395 | 645 |

Table: Distilbert model performance on unaltered in-sample test set

## Footnotes

1. https://huggingface.co/distilbert-base-german-cased [↩]
2. https://www.scripts-berlin.eu/ [↩]
3. https://github.com/nicolaiberk/_rrpviol_med [↩]
4. Note that the paper distinguished three studies. For the sake of simplicity, we clump the first two into one [↩]
5. This test set was of course ot used in training. [↩]
6. the full list of excluded terms can be found in Appendix A of the final report [↩]
7. Assessing data with no blindfolding but temporal restrictions affects the model in a similar way as below, but with higher estimates for AfD and Greens, and lower estimates for FDP. [↩]

## References


Devlin, Jacob, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-training of deep bidirectional transformers for language understanding." *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference* 1 (Mlm): 4171–86. http://arxiv.org/abs/1810.04805.

Gentzkow, Matthew, and Jesse M Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78 (1): 35–71. https://doi.org/10.3982/ecta7195.

Niven, Timothy, and Hung-Yu Kao. 2019. "Probing Neural Network Comprehension of Natural Language Arguments." *CoRR* abs/1907.07355. http://arxiv.org/abs/1907.07355.

Roßteutscher, Sigrid, Harald Schoen, Rüdiger Schmitt-Beck, Bernhard Weßels, Christof Wolf, and Alexander Staudt. 2019. "Rolling Cross-Section Wahlkampfstudie mit Nachwahl-Panelwelle (GLES 2017)." GESIS Datenarchiv, Köln. https://doi.org/ZA6803 Datenfile Version 4.0.1, 10.4232/1.13213.

Widmer, Philine, Elliott Ash, and Sergio Galletta. 2020. "Media Slant is Contagious." *SSRN Electronic Journal*, 1–42. https://doi.org/10.2139/ssrn.3712218.


## Corrections

If you see mistakes or want to suggest changes, please create an issue on the source repository.

## Reuse