

# NLP Project Proposal: Classifying Newspaper Political Bias with Domain-Context augmented Transformers

Tom Arend

t.arend@phd.hertie-school.org

Nicolai Berk

nicolai.berk@gmail.com

## Abstract

*Newspapers are one of the most important institutions influencing political outcomes in contemporary democracies. They have the power to affect voting behaviour [4, 9], as well as polarise the electorate [11] or motivate them to turn out to vote [6]. Therefore, it is surprising that few papers deploy state-of-the-art Deep Learning technologies to classify ideological bias in news articles. We first fine-tune a transformer neural network on party press releases, which we subsequently augment with domain-relevant contextual data, before we go on to estimate ideological bias in news articles. We proceed to compare both models to an ambitious baseline model and prior accuracy scores in the literature.*

## 1. Introduction

Political scientists often want to classify texts into categories based on their content. For example, a researcher might want to classify congressional speeches based on their ideological content, or find out whether political speeches invoke nationalist themes, or even classify news articles based on the narratives they use (Krebs 2015). As many of the issues contained within politicians' speeches are so nuanced that even word choice can convey policy positions, it is unsurprising that most of the efforts are centered around identifying ideological components of speeches, press releases and campaign adverts. Newspapers, or more precisely editorials themselves have received comparatively less attention. Yet they are one of the most influential institutions influencing political outcomes in contemporary democracies.<sup>1</sup>

Newspapers have the power to affect voting behaviour, as well as polarise the electorate or motivate them to turn out to vote. While the impact of political news media has often been studied, their political bias has usually been assumed

a-priori. To improve our understanding of the effects of news medias' ideological bias, it is important to understand how this bias compares across time and space. A number of studies have drawn on NLP and state-of-the-art Deep Learning techniques to automatically classify and identify newspaper slant. However, the existing models tend to (for the most part) ignore domain knowledge inherent to political science to improve the classification of newspaper slant.

We aim to address this shortcoming in the literature by putting forward a transformer neural network augmented with contextual information to beat state-of-the-art classifiers. We aim to achieve this by fine-tuning an existing pre-trained model (either BERT, RoBERTa, DistilBERT or XLNet) on political texts and then using this model to identify newspaper biases. In a first phase, we intend to fine-tune an already pre-trained model on party press releases, using contextual information to supplement it. In a second phase we apply this model to classify newspaper bias. We compare both models to an ambitious baseline model using BERT. Estimates are validated using a number of technical tests and a comparison with previous findings in the literature.

## 2. Related Work

Measuring ideology is the prime use-case of automated text analysis in political science. While some work has used supervised learning to understand political polarisation [1] as well as media slant [16], this existing work estimates the polarisation of political actors or newspapers across time, or simply the similarity of one news source to others. Existing models have for the most part relied on bag-of-words approaches, including the scaling of texts using supervised [10] or unsupervised models [14]. Unsurprisingly, the performance of these models has so far proven unreliable, lagging behind the accuracy of human coders or expert judgement [3, 7, 8]. Recent applications have assessed the ability of deep learning models to replicate human coding of party manifestos [2] or estimate ideology directly using supervised classification [13]. These applications employ

<sup>1</sup>Here is the github repository that we will be using for the project: [https://github.com/nicolaiberk/nlpdl\\_project](https://github.com/nicolaiberk/nlpdl_project)

RNNs or CNNs. We feel they could be improved upon with state-of-the-art transformer neural networks.

Much less work has been devoted to the measurement of the ideological leaning of newspapers. While some work has assessed the impact of endorsements of specific candidates [4, 9], there have been few attempts to scale newspapers on ideology. Gentzkow and Shapiro estimate slant in US newspapers by identifying bi- and trigrams’ indicative of a congressional speakers’ party, and apply the resulting dictionary to newspapers to scale them [6]. More recently, Widmer et al. have assessed polarisation in the US media environment using a supervised model. They train a classifier on bigrams, predicting whether content was produced by a left-leaning network (CNN) or a right-leaning network (Fox news) [16].

We believe both approaches are likely inferior to more complex deep learning models, as such novel approach would not incorporate idiosyncratic phrases used by the specific networks identified to train the data. Using party labels to train classifiers is more straightforward, as it places newspapers within the existing context. However, any algorithm trained on a different dataset than applied to must be subject to careful validation.

### 3. Motivation

To our best knowledge, newer Sequence-to-sequence models have so far not been applied to the classification of newspaper biases. Beyond the implicit motivation inherent to such a research gap, we fundamentally believe that providing a state-of-the-art model to identify ideological biases in newspapers is also beneficial to the wider political science discipline and real-world applications alike.

As stated-above, newspapers are an important institution in the analysis of political phenomena, ranging from polarization to voter turnout. If the existing approaches to classify newspaper slant can be approved upon, or even just complemented, we could provide an additional tool for researchers to study the impact of newspaper slant. A working state-of-the-art model might even renew interest in the subject matter and encourage researchers to find new and exciting applications for it.

Outside of academia, a confident and robust classifier of newspaper bias, might help readers identify when they are reading an article that is overly partisan. This would perhaps encourage them to approach certain news sources with more scepticism and hold news outlets to higher editorial standards. In the long run, the highlighting of biases in articles might counteract the worrying polarization of

entire electorates.

On a more personal note, the chosen project would also align with an area that both authors have a substantive research interest in. It is this combination of domain knowledge with cutting edge technical approaches that provides ample motivation to conduct this project.

### 4. Evaluation

We believe that there will be two separate aspects to the evaluation of the project. The first aspect derives from the personal gain (in an educational sense) that we would like to take away from this project. The second aspect contains the technical robustness and accuracy of the model we train.

1. This concerns the successful training and validation of a state-of-the-art deep learning model for NLP. Since neither of the authors have experience in setting up neural networks, we would consider the baseline for personal success in setting up a valid model that would hold up to the scrutiny of our peers. Evidently, we are more ambitious than that and are hoping to train a number of these models in order to select the most accurate of them. In this way we would hope to gain an advanced level of knowledge of deep learning approaches for NLP that could eventually lead to real and concrete applications in our respective sub-fields of political science.
2. To measure the technical success of our project, we aim to deploy the standard approach in the literature. In order to evaluate the proposed approach, the dataset will be divided into two different subsets: training and validation sets (85%), and a test set (15%) on which we will measure the accuracy of our model. The model may overfit with respect to the validation set. Therefore, a third set, the test set, will be used in order to measure the real performance of the model. Subsequently we will present two statistics that represent the validity of our model; the accuracy rate and the F-measure/test. The current state-of-the-art seems to approach an accuracy of 65-70% [5]. Should our model come close to this score, we would consider the project a success. The best case scenario would be if we could improve on this score by several percentage points.

### 5. Resources

The project will employ the following German language datasets:

1. To fine-tune our transformer model, we intend to employ political texts in the form of party press releases

from Germany that cover the years from 2015-2020. They are labelled with party information and the ideological orientation of that party. This data was collected by the authors for previous projects.

2. The dataset of newspaper articles that will be classified according to their ideological bias, is in the process of being collected. Though we are looking into already existing alternatives. All newspaper articles are scraped from the online archives of the main daily newspapers in Germany.

One of the remaining open questions concerns the language of our datasets. For now, they are only in German. We found contradicting evidence only suggesting that certain mono-lingual pretrained models suffer compared to multilingual pretrained models. We were thus wondering whether we ought to expand the collection of datasets to include English-language political data, such as the e Ideological Books Corpus (IBC), which was compiled by a group of researchers from the University of Maryland [12]. If that were necessary, we could apply our model to a dataset of English-language newsarticles, such as the Media Bias Annotation Dataset [15] or the slant newspaper data by Gentzkow and Shapiro [6].

We will use our individual personal laptops to conduct the training and validation of the model. However, the authors will take advantage of the free computational resources that are available through Google Colab.

## 6. Contributions

Since our group consists of two people we fully intend to divide to workload equally among us. In the following section we will details specifically which task each of us will accomplish.

### Writing Tasks:

- The idea for the project was elaborated in concert in a number of online meetings.
- Whereas Tom did the majority of the writing for the proposal, Nico researched some prior work, set up the group's github account, and collected guides to facilitate the technical collaboration on that platform.
- We expect that Nico will take over most of the writing duties for the midterm progress report.
- The final report will be written in equal parts by both authors. Each will contribute the section on which he worked most. (If Nico conducted the majority of the accuracy tests, he will write the section on validity etc.)

### Computational Tasks:

- We expect to contribute equally to the training and validation of the model.
- For now, both of us have experimented with existing data, but the exact separation of tasks will be established once we have progressed a little further in the course. This will hopefully give us a clearer idea on how to identify and structure the necessary computational tasks to successfully complete our project.

## References

- [1] E. Ash, M. Morelli, and R. Van Weelden. Elections and divisiveness: Theory and evidence. *Journal of Politics*, 79(4):1268–1285, 2017.
- [2] A. Bilbao-Jayo and A. Almeida. Automatic political discourse analysis with multi-scale convolutional neural networks and contextual data. *International Journal of Distributed Sensor Networks*, 14(11), 2018.
- [3] T. Bräuninger, M. Debus, and J. Müller. Estimating hand-and computer-coded policy positions of political actors across countries and time. *Conference of the Midwest Political Science Association*, 2013.
- [4] C. F. Chiang and B. Knight. Media bias and influence: Evidence from newspaper endorsements. *Review of Economic Studies*, 78(3):795–820, 2011.
- [5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. 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):4171–4186, 2019.
- [6] M. Gentzkow and J. M. Shapiro. What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1):35–71, 2010.
- [7] F. Hjorth, R. Klemmensen, S. Hobolt, M. E. Hansen, and P. Kurrild-Klitgaard. Computers, coders, and voters: Comparing automated methods for estimating party positions. *Research and Politics*, 2(2), 2015.
- [8] J. Koljonen, V. Isotalo, P. Ahonen, and M. Mattila. Comparing computational and non-computational methods in party position estimation: Finland, 2003–2019. *Party Politics*, (May), 2020.
- [9] J. M. D. Ladd and G. S. Lenz. Exploiting a rare communication shift to document the persuasive power of the news media. *American Journal of Political Science*, 53(2):394–410, 2009.
- [10] M. Laver, K. Benoit, and J. Garry. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331, 2003.
- [11] Y. Lelkes, G. Sood, and S. Iyengar. The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect. *American Journal of Political Science*, 61(1):5–20, 2017.

- [12] Y. Sim, B. D. L. Acree, J. H. Gross, and N. A. Smith. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 91–101, Seattle, Washington, USA, Oct. 2013. Association for Computational Linguistics.
- [13] A. Simoes and M. Del Mar Castaños. Fine-Tuned BERT for the Detection of Political Ideology Stanford CS224N Custom Project. 6(2017):1–5, 2020.
- [14] J. B. Slapin and S. O. Proksch. A scaling model for estimating time-series party positions from Texts. *American Journal of Political Science*, 52(3):705–722, 2008.
- [15] T. Spinde, L. Rudnitckaia, K. Sinha, F. Hamborg, , B. Gipp, and K. Donnay. Mbic – a media bias annotation dataset including annotator characteristics. In *Proceedings of the iConference 2021*, March 2021.
- [16] P. Widmer, E. Ash, and S. Galletta. Media Slant is Contagious. *SSRN Electronic Journal*, pages 1–42, 2020.