

Institute for Informatics

Ludwig Maximilian University of Munich

# Identification of Political Leaning in German News Articles

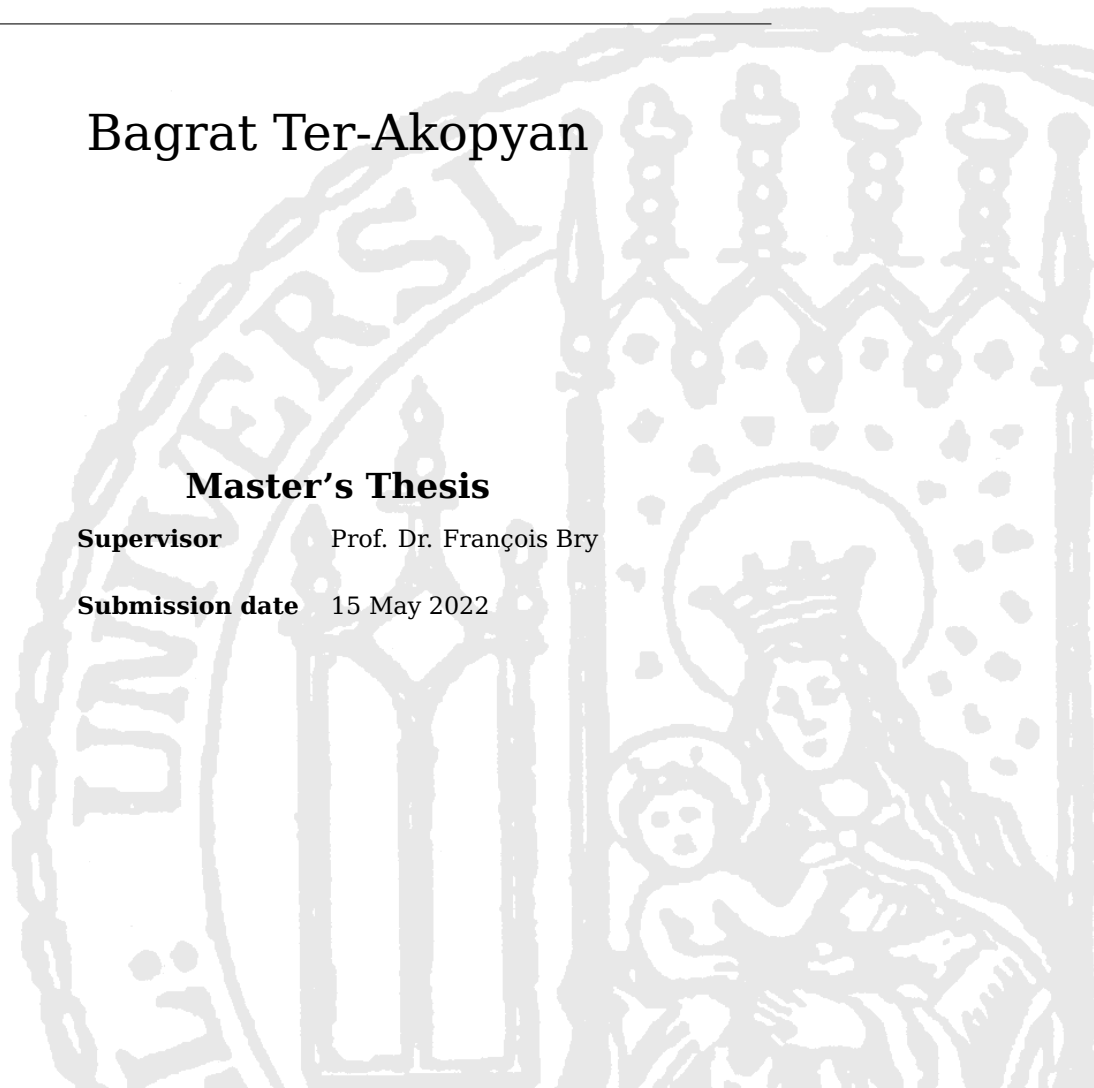
---

Bagrat Ter-Akopyan

## **Master's Thesis**

**Supervisor** Prof. Dr. François Bry

**Submission date** 15 May 2022





---

## Declaration

---

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Munich, 15 May 2022

Bagrat Ter-Akopyan



Media coverage plays an important role in shaping personal and public opinion, as it is often a primary source of information. Especially in today’s world where ideological division with subsequent polarization of the political discourse is constantly increasing across social networks, the responsibility of media professional to ensure an unbiased reporting is growing. However, the coverage often exhibits a specific political leaning that is reflected in the articles and is commonly referred to as *media bias*. Many works already exist in computational social science that explore this phenomenon using English-language newspaper articles. This thesis aims to investigate the phenomenon in the German media landscape. This work uses a new methodology to attribute newspaper articles to a political leaning based on the party affiliation of the politician who shared the article on social media. We used AbgeordnetenWatch — a platform containing short profiles of politicians — to generate a list of German MPs including their Twitter profiles. We extract their tweets, using the Twitter API, and select all tweets that reference an article in the German media landscape. We extract all articles, including the title, the summary, the actual text and several meta information. In the end, we obtain a dataset with a total of over 40 thousand articles from the most popular German media houses and the respective politician who shared the article. We examine the resulting dataset with respect to three research questions: (1) How effective are standard classification approaches with frequency-based features and the standard neural approach in detecting bias from the right-wing political fringe, when using the articles shared by all other parties as negative samples? (2) How effective are the same standard approaches in detecting bias from right-wing political fringe, when training against the articles shared by other parties separately? (3) How do the same approaches perform when applying to the detection of the left-wing party *DIE LINKE*, i.e.: a) when using the articles shared by all other parties as negative samples, analogous to (1); b) when training articles of *DIE LINKE* separately against articles of every other party, analogous to (2). The results of the experiments regarding the second and the third research questions are promising: the state-of-the-art model, BERT, achieves an  $F_1$ -Score between 0.68 and 0.75 in detecting articles shared by the *AfD*, and an  $F_1$ -Score between 0.59 and 0.72 in detecting articles shared by *DIE LINKE*.



---

## Zusammenfassung

---

Die Medienberichterstattung spielt eine wichtige Rolle bei der persönlichen und öffentlichen Meinungsbildung, da sie oft eine primäre Informationsquelle darstellt. Insbesondere in der heutigen Welt, in der die ideologische Spaltung und die daraus resultierende Polarisierung des politischen Diskurses in den sozialen Netzwerken ständig zunimmt, wächst die Verantwortung der Medienschaffenden, eine unvoreingenommene Berichterstattung zu gewährleisten. Die Berichterstattung weist jedoch häufig eine bestimmte politische Ausrichtung auf, die sich in den Artikeln widerspiegelt und gemeinhin als „Medienvoreingenommenheit“ bezeichnet wird. In den Computer-Sozialwissenschaften gibt es bereits viele Arbeiten, die dieses Phänomen anhand englischsprachiger Zeitungsartikel untersuchen. Ziel dieser Arbeit ist es, das Phänomen in der deutschen Medienlandschaft zu untersuchen. Diese Arbeit verwendet eine neue Methode, um Zeitungsartikel einer politischen Richtung zuzuordnen, basierend auf der Parteizugehörigkeit des Politikers, der den Artikel in den sozialen Medien geteilt hat. Wir haben AbgeordnetenWatch — eine Plattform mit Kurzprofilen von Politikern — verwendet, um eine Liste deutscher Abgeordneter inklusive ihrer Twitter-Profile zu erstellen. Wir extrahieren ihre Tweets mithilfe der Twitter API und wählen alle Tweets aus, die auf einen Artikel in der deutschen Medienlandschaft verweisen. Wir extrahieren die Artikel, einschließlich des Titels, der Zusammenfassung, des Textes und Metainformationen. Am Ende erhalten wir einen Datensatz mit insgesamt über 40 Tausend Artikeln der populärsten deutschen Medienhäuser und den jeweiligen Politikern, die den Artikel geteilt haben. Wir untersuchen den resultierenden Datensatz im Hinblick auf drei Forschungsfragen: (1) Wie effektiv sind Standardklassifizierungsansätze mit frequenzbasierten Merkmalen und der neuronale Standardansatz bei der Erkennung von Verzerrungen durch den rechten politischen Rand, wenn die von allen anderen Parteien geteilten Artikel als Negativ-Beispiele verwendet werden; (2) Wie effektiv sind dieselben Standardansätze bei der Erkennung von Verzerrungen durch den rechten politischen Rand, wenn sie mit den von anderen Parteien geteilten Artikeln separat trainiert werden?; (3) Wie schneiden dieselben Ansätze bei der Erkennung der linken Partei *DIE LINKE* ab, d. h.: a) wenn die Artikel, die von allen anderen Parteien geteilt werden, als negative Stichproben verwendet werden, analog zu (1); b) wenn Artikel von *DIE LINKE* separat gegen Artikel jeder anderen Partei trainiert werden, analog zu (2). Die Ergebnisse der Experimente zur zweiten und dritten Forschungsfrage sind vielversprechend: Das State-of-the-Art Modell BERT erreicht bei der Erkennung von Artikeln, die von *AfD* geteilt werden, einen  $F_1$ -Score zwischen 0,68 und 0,75 und bei der Erkennung von Artikeln, die von *DIE LINKE* geteilt werden, einen  $F_1$ -Score zwischen 0,59 und 0,72.





---

## Acknowledgments

---

Foremost, I would like to thank Prof. Dr. François Bry, who supervised and reviewed my master thesis. For the helpful suggestions, the constructive criticism and motivating words during this extraordinary time, I would like to thank him very much.

Furthermore, I would like to thank my parents, Marina and Arkady, who supported and encouraged me at all times. Last but not least, I would like to thank my “fellow sufferers”, Raphael and Giuliano, who always stood by me with advice, support and an open ear.



---

## Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background: Media Bias</b>	<b>5</b>
2.1	The Nature of Bias . . . . .	5
2.2	Causes and Forms of Media Bias . . . . .	6
2.3	Effects of Media Bias . . . . .	7
<b>3</b>	<b>Related Work</b>	<b>9</b>
3.1	Detecting Media Bias . . . . .	9
3.2	Analyzing Media Bias . . . . .	10
3.3	Modifying Media Bias . . . . .	11
3.4	Building Media Bias Datasets . . . . .	12
<b>4</b>	<b>Corpus Construction</b>	<b>15</b>
4.1	Collecting Members of Parliaments . . . . .	16
4.2	Collecting Tweets . . . . .	18
4.3	Collecting News Articles . . . . .	21
4.4	Dataset Statistics . . . . .	22
<b>5</b>	<b>Experiments</b>	<b>25</b>
5.1	Research Questions . . . . .	26
5.2	Experiment Settings . . . . .	26
<b>6</b>	<b>Evaluation</b>	<b>33</b>
<b>7</b>	<b>Conclusion</b>	<b>39</b>
7.1	Discussion . . . . .	40
7.2	Outlook . . . . .	41



# CHAPTER 1

---

## Introduction

---

The diversity of opinions and beliefs in society is what makes our democracy most effective and stable. Questioning one's views, solidifying them, or even rejecting them in case of doubt is an existential part of human life. These processes are favored when individuals are faced with information that does not correspond to their pre-existing views or beliefs (Gentzkow and Shapiro, 2011).

Internet access dramatically reduces the cost of acquiring information from a wide range of sources. Thus increases people's ability to gain knowledge, form their own opinions, and access socially relevant topics. This unlimited access to unbiased information is essential for shaping a balanced view on realities. News articles play a fundamental role in shaping personal and public opinion, as they often serve as an important source of information. According to De Saussure (2011) languages are a system of signs — "pairing of form and meaning"<sup>1</sup> — and so news coverage is more than just a reporting of plain facts. By putting facts into wider context, news articles often convey a specific point of view. As a result, the way journalists cover a topic can have a profound influence on our decisions. Bias tonality, misleading word choice, and other expressions of media bias can have an undesirable influence on how individuals perceive collective topics.

### Motivation

As mentioned above, unlimited access to a growing number of information sources is essential in the sense that every individual has the chance to participate in reporting that reflects the entire spectrum of opinion, including views that do not correspond to his or her own. But on the other hand, it carries the danger of ideological self-segregation for consumers, limiting themselves to sources that are expected to support their prior point of view (Mullainathan and Shleifer, 2005). Sunstein (2001) describes: "people restrict themselves to their own points of view – liberals watching and reading mostly or only liberals; moderates, moderates; conservatives, conservatives; Neo-Nazis, Neo-Nazis". And in fact, the ideological self-segregation followed by the polarization of online politi-

---

<sup>1</sup>Bender et al. (2021)

cal discourse on Twitter (Yardi and Boyd, 2010; Conover et al., 2011), Facebook (Bakshy et al., 2015), and Reddit (An et al., 2019) have gained increasing attention in the computational social sciences and the natural language processing community, short *NLP*, over the last years.

This development increases the responsibility of the media in reporting news as unbiased as possible. In 2019 the International Federation of Journalists revised "The Global Charter of Ethics for Journalists", article ten states that falsifying a fact violates ethical principles of journalism: "The journalist will consider serious professional misconduct to be plagiarism, distortion of facts, slander, libel, defamation, unfounded accusations"<sup>2</sup>. However, news reporting often reveals an internally intended bias that is reflected in the articles and commonly addressed to as *media bias*. There exist several factors that can influence this bias: outlet ownership or the source of income of the media outlet, as well as particular political or ideological stance either of the outlet itself or its audiences (University of Michigan, 2014). Media coverage can reveal bias in several ways. For instance, the presentation of news articles within a newspaper or on its website can differ in various ways: while some articles are small and placed at the very bottom or end of a newspaper, others are at the very top in huge, manipulating the attention the article will receive from the readers (Bucher and Schumacher, 2006). Field et al. (2018) differentiate between two other concepts in this context:

**Agenda-Setting:** before publishing a story, journalists select *events* based on the self-determined relevance and the corresponding *source*. Often, journalists do not cover the whole topic or event at once, but select certain *information* that they finally publish in the news article.

**Framing:** to steer a reader's opinion on a specific topic in a certain direction, journalists often use intended *word choices* with either a positive or negative connotation towards an entity.

Additionally, due to the fact, that readers tend to "follow" news that are congruent with their pre-existing views and beliefs (Sunstein, 2001; Milyo and Groseclose, 2005) social media amplifies the impact of biased reporting even more. In the literature, this effect on social media is often referred to as "echo chamber" or "filter bubble", where consumers just reinforce their internal biases. Furthermore, due to the consequences of the information overload, regional and linguistic affiliation, or personal interests most newsreaders often tend to consume only a fraction of available news outlets.

Motivated by these problems and its huge impact on the society, the automated identification of media bias, and the analysis of news articles in general, has moved more and more into the focus of computer science research. Since the work of Lin et al. (2006), much research has been done in three different directions in the context of political bias: *analyzing*, *modifying* and *detecting* the media bias. Since most solution approaches for detecting media bias are based on machine learning methods, the creation of a suitable dataset is a fundamental prerequisite. To the best of our knowledge, the standard approaches in detecting political article-level bias achieve weak results. Even if using second-order bias information, i.e., sentence-level bias (Chen et al., 2020a), outperforms the best existing approach so far, it still does not achieve break through results on this task.

---

<sup>2</sup>International Federation of Journalists, 2019, <https://www.ifj.org/who/rules-and-policy/global-charter-of-ethics-for-journalists.html>, accessed: 2022-02-23

The weaknesses and the poor performance of existing approaches are mainly based on two aspects. Most of the existing approaches to analyzing and, consequently, detecting media bias disregard the insights from the social sciences. The resulting models in computer science are mostly very simplified and do not bring any new insights compared to the models and findings from the social sciences (Hamborg et al., 2018). The second aspect concerns the assumptions made for the creation of datasets to train bias detection models:

A1: **Raters' bias:** *Ratings of political content are independent of political leaning of the raters* (Gentzkow and Shapiro, 2010)

A2: **Media-level bias and article-level bias:** *News articles correspond to the political leaning of their source outlet* (Potthast et al., 2018; Kulshrestha et al., 2018)

A3: **Topic-level bias:** *Political leanings of news outlets stay stable across different topics* (Groseclose and Milyo, 2005; Bakshy et al., 2015; Kulshrestha et al., 2017)

Ganguly et al. (2020) point out that these assumptions do not always hold, which could be instrumental in poor performance of the models trained on them.

## Methodology

Considering the weaknesses of existing approaches to construct a labeled dataset (Ganguly et al., 2020) and the idea to infer the political leaning of individual news outlets based on selective sharing of parliaments members (Freitag et al., 2021), this thesis provides a novel approach for constructing such a dataset, labeling articles according to the party affiliation of the politician who shared the corresponding article. Regarding the method to create the dataset, we define three research questions:

- Q1.** How effective are standard classification approaches with frequency-based features and the standard neural approach in detecting bias from the right-wing political fringe, when using the articles shared by all other parties as negative samples?
- Q2.** How effective are the same standard approaches in detecting bias from right-wing political fringe, when training against the articles shared by other parties separately? As in the first research question, we label articles shared by *AfD* politicians as biased and the others as unbiased.
- Q3.** How do the same approaches perform when applying to the detection of the left-wing party *DIE LINKE*, i.e.: a) when using the articles shared by all other parties as negative samples, analogous to **Q1**; b) when training articles of *DIE LINKE* separately against articles of every other party, analogous to **Q2**.

The data collection process consists of multiple steps: (1) Collecting member of German parliaments; (2) Collecting Tweets for every politician; (3) Extracting news articles from corresponding webpages of German news outlets. To study the defined research questions, we deploy standard feature based models, Logistic Regression, Linear SVM and Naive Bayes, trained on TF-IDF vectors of excerpts' n-grams(1–3), and a pretrained state-of-the-art BERT model, that we further fine-tune on the excerpts of news articles regarding the defined classification tasks.

## Contribution

In summary, the contribution of this thesis is three-fold: (1) To the best of our knowledge, this thesis provides the first fully automated and scalable approach to build a political media bias dataset with over 40 thousand labeled news articles, inferring the labels of news articles from the political affiliation of the politician, following the hypothesis that social media users share and follow the content that corresponds to one's own values and beliefs (Stefanov et al., 2020; An et al., 2012; Morgan et al., 2013; Ribeiro et al., 2018; Freitag et al., 2021); (2) Since most of the research work with English datasets, we consider the construction of a German dataset as the second contribution; (3) Based on the results of the experiments regarding the second and the third research question, we show, that it is possible to infer the labels of news articles from the political affiliation of the politician.

The remainder of this thesis is structured as follows: Chapter 2 gives a brief introduction to the theoretical background knowledge for better understanding of *media bias*, and discusses the existing approaches for analyzing media bias from social science perspectives. Chapter 3 provides an overview on existing and related approaches for building political media bias datasets and discusses their strength and weaknesses, as well as an overview on existing research in the most common downstream tasks in computer science in context of political bias. Based on the weaknesses of existing approaches to create a political media bias dataset, Chapter 4 describes the approach developed in this thesis to create a suitable corpus, followed by presenting the descriptive statistics and properties of the final dataset. Adapted from related work, Chapter 5 presents the standard machine learning approaches to predict the political media bias, as well as the neural state-of-the-art approach in NLP. The evaluation of the experiments is discussed in Chapter 6, followed by a brief summary and conclusion in Chapter 7.



---

### Background: Media Bias

---

This chapter provides an overview of the *media bias* phenomenon. First, we explain the different definitions and characteristics of *media bias* that have become established in research over time. We then describe the different causes and forms of *media bias*. We then show what effect media bias has on society.

### 2.1 The Nature of Bias

*Media bias* has been studied in social sciences at least since the work of White (1950). Over time, several definitions have become established in research. Basically, literature differentiates between *intentional* and *unintentional bias* (Hamborg et al., 2018). According to the classical definition of Williams (1975), media bias can only be considered as such in the presence of certain properties: (1) it must be volitional, or willful, i.e., it must reflect a conscious action or decision; (2) it must be influential, otherwise it is irrelevant; (3) it must pose a threat to widespread conventions, lest it be dismissed as mere “crankiness”; and (4) it must be persistent, rather than one-shot. In contrast, on the one hand the news value (Harcup and O’neill, 2001) and on the other hand the perception of the readers due to different backgrounds (Oelke et al., 2012) can trigger the *unintentional bias*. This thesis focuses exclusively on the media bias according to the first mentioned definition.

In the social science literature, there exist other definitions of media bias, as well as their specific manifestations. Mullainathan and Shleifer (2002) draw a distinction between two basic types of biased reporting. First, they consider the traditional left or right bias as *ideology*, i.e., preference to report news from one political side. Second, they define bias created from need to tell a memorable story as *spin*. Although the motivations of the respective bias expressions are fundamentally different, the accuracy of news reporting is unfortunately the same in both cases. Another commonly used definition makes a distinction between the following three types of bias: *gatekeeping*, *coverage* and *statement* (D’Alessio and Allen, 2006). Gatekeeping bias, also referred to as selection (Groeling, 2013) or agenda bias, arises during the act of simplification: selection of events to report. This process consequently leads to information that is

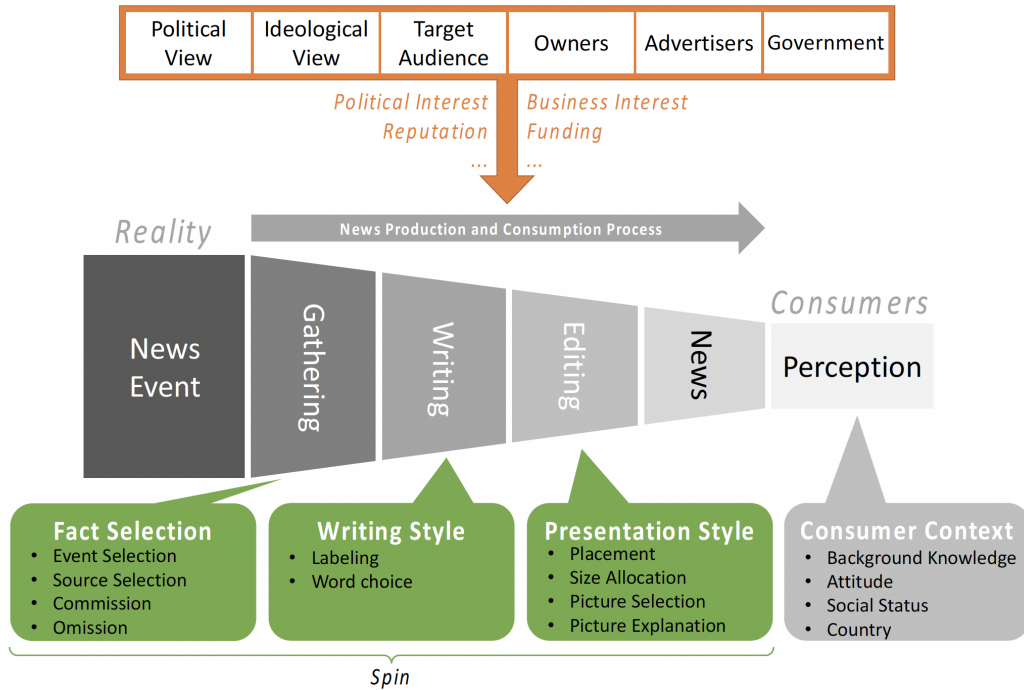


Figure 2.1: Causes and forms of media bias (Hamborg et al., 2018)

necessarily discarded. Coverage bias describes the physical discrepancy in reporting between two sides regarding an issue. This disbalance is often measured in column inches for newspapers and newsmagazines. Statement bias occurs when journalists include their own opinions while reporting on a factual issue. Gentzkow et al. (2015) divides bias in *filtering* and *distortion*. With filtering, journalists provide their readers with a one-sided selection or summary of all available information. This can deliberately steer them in a political direction. Puglisi and Snyder Jr (2015) refers to the phenomena as *partisan filtering*. Distortion occurs when the reporting information deviates from the reality. Whereas selection bias reduces the information or events to cover, *presentation bias* distorts the content of the stories (Groeling, 2013).

## 2.2 Causes and Forms of Media Bias

To better understand the different forms of media bias and their possible specifications, it is necessary to understand the process of news production in detail. The different forms of embodiment of biases described by Baker et al. (1994) are mapped to stages (gray) in Figure 2.1. Different aspects can have a direct or indirect effect on the process of news production. The aspects mentioned in Figure 2.1 in the orange rectangle point to intrinsic and extrinsic motives behind the media bias. The explanation of the individual phases and the corresponding causes and forms of media bias are based on the literature review of Hamborg et al. (2018).

**Internal motives:** The internal *political* and *ideological* views of the media outlets are present at all levels of the organizations, i.e., overall at the company level as well as at the personal level of each journalist, encoding their own political orientation in

the coverage (Groseclose and Milyo, 2005). Baron (2006) also notes that journalists often resort to polarizing and biased wording when it can lead to their own professional advantage.

**External motives:** Journalists often portray the news in line with the views of the news outlet's *target audience* (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010). Conversely, readers often change their news sources if the preferred ones too often holds a contrary opinion (Sunstein, 2001; Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2010). *Owners* and *advertisers* of the news outlets themselves are often the cause for biased reporting. In case they are involved in a conflict of public interest, journalists often avoid reporting on that topic in their favor (De Vreese, 2005; Gilens and Hertzman, 2000). A similar problem exists with the government. Because journalists often depend on information from within the government, they can avoid negative reporting here as well (Herman, 2000; D'Angelo, 2018).

Internal as well as external motives can be decisive for biased reporting at all stages of news production process (gray funnel in Figure 2.1). In the following paragraphs, we explain the process and the corresponding aspects leading to bias.

**Gathering:** *Gathering* refers to the process of fact selection. Regardless of the motives, the relevance of all events happened is not equally distributed. Thus, journalists first start with the *event selection*. Subsequently, journalists have to choose their sources of information, e.g., press releases of news agencies, other newspapers, personal experience reports. Due to the fact, that each story is covered in different scope, journalists often have to choose the aspects of the story to cover and which to ignore. This process refers to the step of *commission* or *omission*.

**Writing:** In the *writing* stage, journalists can articulate deliberate bias by employing different *writing styles* through *labeling* and *word choice*. *Labeling* is about giving an event, action, or attribute, a positive, none, or even a negative connotation, while a specific *word choice* can describe one and the same unit with a different interpretation.

**Editing:** Furthermore, in the *editing* phase, biased views are particularly emphasized by visual aspects. Placement, the size, as well as the image selection and its explanation also play a significant role in public perception.

In summary, there exist various sources of bias across all stages of the news production process. Ultimately, in the *perception stage*, even here can still come to the reinforcement of the bias in one direction or another due to different *consumer context*. The perception of the information always strongly depends on personal characteristics of the respective reader, such as, *background knowledge*, *attitude* to a given topic, the effect of the story on their *social status*, or their *country*.

## 2.3 Effects of Media Bias

The different manifestations of media bias, described above, affect public perception in all areas of public life, and thereby influence the political decisions of both citizens and decision-makers in politics (Bernhardt et al., 2008). Despite the rise of social media,

and development of new content formats, newspapers remain one of the most important news sources for publicity. According to "Deutschland-Portal," an independent service provided by "Fazit Communication GmbH" in cooperation with the German Foreign Office, German-language printed newspapers have a reach of 56 percent of the German-speaking population. Extending the view to the digital edition of the newspapers, it is even 84.6 percent<sup>1</sup>. Therefore, biased and unbalanced reporting leads to a polarized society. Nowadays, social media, e.g., Twitter, Facebook or Instagram, even amplify and accelerate the effect of polarization. The increased polarization is due to a phenomenon referenced in science as homophily, i.e., the axiom that similarity breeds connection (Chun, 2018). As a result, readers only "follow" and "share" news that is consistent with their pre-existing opinions (Groseclose and Milyo, 2005; Sunstein, 2001; Mullainathan and Shleifer, 2002), "making cyberspaces a series of *echo chambers*" (Chun, 2018). The strong polarization leads to a divided society, which in turn can have an impact on election results (Druckman and Parkin, 2005) and, lead to disunity on contentious issues.

According to Scheufele (2000); Druckman and Parkin (2005), there exist three ways how biased coverage affects the perception: *agenda setting*, *priming* and *framing*, whereby priming and framing are seen as an extension of agenda setting. Priming theory points out that the prior coverage on an issue is highly significant for the consumer's evaluation of the particular topic. Agenda setting is comparable with the first stage — *gathering* — of the news production process, where journalists consciously select the topics to be reported on in order to direct their readership's attention in a particular direction. Furthermore, journalists can portray a topic from different perspectives. This technique is called *framing* and allows, with regard to a fact, "to promote a particular interpretation" Entman (2007).

---

<sup>1</sup><https://www.deutschland.de/en/topic/culture/media-in-germany-user-figures>, accessed: 2022-03-01

The work of Lin et al. (2006) was the first to analyze the phenomenon of *media bias* in computer science. Since then, *media bias* was addressed under different terms, e.g., *perspective* (Lin et al., 2006), *ideology* (Iyyer et al., 2014), *truthfulness* (Rashkin et al., 2017), and *hyperpartisanship* (Kiesel et al., 2019). The research branches and the corresponding tasks are manifold. The following section gives a brief introduction to different research directions in context of media bias, including media bias analysis, media bias detection, bias flipping and constructing political media bias datasets for each task respectively.

### 3.1 Detecting Media Bias

To the best of our knowledge, the work of Lin et al. (2006) is the first that studied media bias in the area of computer science. In their work, they addressed whether computers can identify the *perspective* of a document. In this context, they define *perspective* as a "subjective evaluation of relative significance, a point-of-view"<sup>1</sup>, referring to a related concept such as media bias. They considered the problem of learning individual perspectives as a classification task and developed statistical framework models to analyze how perspectives are manifested in word usage. They evaluated their models on articles about Israeli-Palestinian, i.e., topic specific, conflict at the document and sentence levels. The results of the work reveal that statistical models can classify the perspective of a document with high performance when the whole document collection is related to a concrete topic. The weaknesses of such models are obvious: they cannot be generalized to other topics and have to be trained anew for each topic.

While the majority of research is dedicated towards the *lexical bias* captured by linguistic attributes such as word choice and syntax, the work of Fan et al. (2019) studies the effects of *informational bias*, i.e., reporting selective content to manipulate reader's opinion. For that purpose they create a new dataset, **BASIL**, of 300 news articles annotated with 1,727 bias spans. They show with evidence that informational bias occurs

---

<sup>1</sup>The American Heritage Dictionary of the English Language, 4th ed.

more frequently in news articles than lexical bias. The paper further analyzes how information bias manifests differently in news articles depending on the publisher. For informational bias prediction, they fine tune BERT (Devlin et al., 2019) on the labeled data as baseline model. The findings of the work are manifold. First, they show that *lexical bias* often occurs at the beginning, while *informational bias* is used more frequently than the lexical and is distributed uniformly across the entire article. Second, nearly half of the informational bias comes from quotes. It reveals a bias strategy in which publishers intentionally select quotes reflecting their own opinions.

Based on the knowledge that most feature-based and neural text classification approaches relying only on the distribution of low-level lexical information fail to achieve high performance in detecting media bias, another work of Chen et al. (2020a), inspired by Wachsmuth et al. (2015), study how second-order information about biased statements in an article can help in improving the performance of a detection model. In detail, they make use of the probability distributions of the frequency, positions, and sequential order of lexical and informational sentence-level bias in a Gaussian Mixture Model. On the **BASIL** data set, provided by Fan et al. (2019), they show that frequency and positions of biased statements strongly impact article-level bias and that standard models for sentence-level bias detection using second-order information obviously outperforms those without.

Spinde et al. (2021b) developed an automated system to identify bias inducing words based on linguistic and context-oriented features. Due to the lack of large-scale gold-standard data sets, they designed a prototypical and diverse data set for this purpose. Relying on media bias ratings of <https://allsides.com>, they handpicked 1,700 sentences from around 1,000 articles. They use crowdsource annotators from Amazon Mechanical Turk for final annotations on word level. Unlike similar research in that field, the constructed dataset pay attention to background information on the participants’ demographics, their ideology, and opinion about media in general, increasing the transparency and reliability. In contrast to deep learning techniques, the approached feature-oriented system allows for descriptive analysis and the ability to interpret the results. Furthermore, they make out linguistic, lexical, and syntactic features that in the end can be used as indicators for media bias detection.

While most of the research work with English datasets, Spinde et al. (2020) propose a method for analyzing media bias in German coverage about refugee crisis. The approach combines three different components: an IDF-based component, a specially created bias lexicon, and a linguistic lexicon to detect bias on the word level. For the analysis, they collected news articles from four German news outlets, *Süddeutsche Zeitung*, *TAZ*, *Südkurier*, and *BILD*. Using the data with the collection of articles provided by Bojar et al. (2014)<sup>2</sup>, allows more accurate training of word embeddings. Still, the evaluation of the provided methodology is strongly dependent on manually annotations on the word-level to create a ground-truth dataset.

## 3.2 Analyzing Media Bias

Based on Wikipedia’s revision history, Recasens et al. (2013) make use of edits that are associated with *neutral point of view (NPOV)*<sup>3</sup> tags. This policy consists of a collection of principles, which includes “avoiding stating opinions as facts” and “preferring nonjudgmental language”. Following this approach, Recasens et al. (2013) construct a

<sup>2</sup>Contains about 90M sentences from over 40 sources.

<sup>3</sup>[https://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view), accessed: 2022-03-24.

dataset consisting of sentence pairs, before the edit in the biased and after the edit in the corresponding unbiased form. The dataset allows for a more detailed analysis of the linguistic implementation and nature of bias. According to their analysis, they identify two classes of edits: **framing** and **epistemological** bias, which allows creating a bias lexicon, which is then in turn fed into a classifier to predict these bias-inducing words.

Lim et al. (2018) concentrate on understanding the underlying nature of bias and its manifestation, as well as creating a robust gold standard dataset on word and sentence levels with the help of crowdsourcing. The analysis of the characteristics of the user annotations reveals that identifying bias-induced words is strongly subjective and depends on annotators personal background, making the agreement of all readers very difficult. Further study of discriminative characteristics of biased text suggests that linguistic features, e.g., sentiment words, seem to be a good indicator for bias.

Chen et al. (2020b) compile a dataset based on *allsides.com*<sup>4</sup> and adds additional labels based on information provided by *adfontesmedia.com*<sup>5</sup>. In their work, the authors introduce three additional bias categories on the level of news portals. First, *political bias*, e.g., *neutral* if it is labeled as "skew left/right" or "neutral", or *political biased* in case it is labeled as "most extreme left/right" or "hyperpartisan left/right". Second, they define the *unfairness*. A portal with one of the labels "original fact reporting", "mix of fact reporting and analysis", "analysis", or "opinion" are defined *fair*, while portals with labels "selective story", "propaganda", or "fabricated info" are defined as *unfair*. Third, portals that are politically unbiased and fair are considered as *objective*, otherwise as *non-objective*. The main aspect of the study is, to analyze the bias at different text levels of text granularity. With training sequential models for bias detection and applying a reverse feature analysis, they achieve good results in revealing the granularity level of bias.

### 3.3 Modifying Media Bias

Chen et al. (2018) introduced another new task in the context of media bias: "flipping" the bias of news articles. Given a bias (left or right), rewrite the article to have an opposite bias while keeping the topic. Also in this work, the data corpus is created based on bias-labeled articles from *allsides.com*. The findings of the comparison of biased and sentimental texts reveals that the set of discriminative words of biased texts differ from those of sentimental, and that some bias occurs only at paragraph or article level. The results of a proposed cross-aligned autoencoder to rewrite article headlines suggest that current state-of-the-art concepts have troubles regarding this task.

A slightly different task is addressed in the work of Pryzant et al. (2020). Instead of "flipping" the bias, they developed an approach to "neutralize" biased text, i.e., automatically rewrite biased text into a more objective factual form. The dataset of 180,000 sentence pairs were taken from edits from Wikipedia that modified several framing, assumptions, and slants from sentences. They use the same method to identify bias as Recasens et al. (2013): using Wikipedia's *neutral point of view (NPOV)* policy. The paper propose two encoder-decoder architectures as baselines. First, considering the biased source  $\mathbf{s}$  the *CONCURRENT* system identify problematic words and generates a neutralized version  $\hat{\mathbf{t}}$  in one step. This model is easy to train and apply, but has some limitations in the sense of interpretability and controllability. Second, the interpretable

<sup>4</sup>A news aggregator collecting news article about American politics from all sides of political spectrum, categorizing them as *left*, *center* or *right*.

<sup>5</sup>A portal quantifying media bias of US news publishers.

*MODULAR* system address the process in two separate stages: (1) the **Detection Module** for identifying biased text and (2) **Editing Module** that takes the biased source sentence and rewrite it into a more neutral form.

### 3.4 Building Media Bias Datasets

Most of the research described above show some drawbacks in creating a suitable dataset for addressing the task, respectively. Following limitations of the datasets were recognized by Hamborg et al. (2018): (1) topic limitation (Lim et al., 2018, 2020), (2) exclusively focus on framing (Baumer et al., 2015; Hamborg et al., 2019), (3) target-oriented annotations (Hamborg et al., 2019; Fan et al., 2019) or (4) datasets are basically too small, which is the case for almost all data sets. One of the main difficulties in current research on media bias detection is the lack of representative and diverse datasets with annotations of bias on different level of granularity: word, sentence, paragraph or the entire article. The background of annotators represents important information regarding their annotations. Most of the work described above do not take into account that important aspect. To overcome that lack, Spinde et al. (2021c) present a matrix-based methodology and provide a self-developed annotation platform to collect such data. Furthermore, they present **MBIC**<sup>6</sup> dataset containing 1,700 statements with several types of media bias instances. The available information on annotator characteristics and their individual background helps significantly to understand the perception of media bias, that in turn can lead to significant improvement in detecting bias.

Spinde et al. (2021a) identify another problem regarding the datasets: While most of the research create their datasets based on questioning the perception of bias either students, experts, or crowdsource workers, almost none of them accurately report the process of survey evaluation or the selection of questions. This lack of agreement regarding the process, partial questions overlap across several studies, and diversity in methods and definitions result in limited comparability on media bias perception between studies. Thus, the focus of the work was to develop a failsafe question catalog for further research.

Ganguly et al. (2020) evaluate the weaknesses of common assumptions made in research to create a Political Media Bias Dataset. They identify three basic assumptions:

- (1) **Raters' bias**, i.e., raters' political leanings do not affect labeling tasks
- (2) **Media-level and article-level bias**, i.e., news articles follow their source outlet's political leaning
- (3) **Topic-level bias**, i.e., political leaning of news outlets is stable across different topics

For the purpose of the evaluation, Ganguly et al. (2020) collect news articles from U.S. outlets representing different political orientation on controversial topics like "Gun policy" and "Immigration" over a period of 3-month in 2018. With the help of Amazon Mechanical Turk, they build a gold labeled dataset with manual annotation on article-level. The findings of the paper reveal that even with small datasets, the three common assumptions could be invalid.

---

<sup>6</sup>Media Bias Including Characteristics.



**Summary** In this chapter we have shown, that there already exist a very wide range of research done in context of media bias. While most research is specifically and intensively concerned with the question of identification and analysis of linguistic features of bias, and primarily specializes in technical approaches, others try to neutralize or “flip” the bias to the opposite, which can be seen as a downstream task of text generation. Other work, in turn, addresses conceptual issues such as the perception of bias and the design of comparable questionnaires and reviews of the assumptions made to label bias as such. The aforementioned weaknesses, especially from the work of Ganguly et al. (2020) regarding the three common assumptions, inspired us to develop a scalable and automated methodology to derive the political leaning of newspaper articles, being completely independent of any kind of surveys. The following Chapter 4 describes the process of building our Political Media Bias Dataset.



---

## Corpus Construction

---

Crucial to the success of all the research fields described in Chapter 3, and of all machine learning tasks in general, is foremost the quality of the data. Considering the indicated weaknesses described in Section 3.4, this chapter describes the methodology used to create the appropriate dataset for the classification of media bias. Figure 4.1 illustrates the intermediate steps of the dataset construction. Based on the assumption that people on social networks, including Twitter, share significantly more content that corresponds to their own worldview and follows their political leaning, we tag the articles according to the political orientation of the particular person who shared the content. Section 4.1 describes the process how we retrieve the German politicians, including their Twitter profile names. In Section 4.2 we describe the interaction with the Twitter API to collect the tweets of the previously selected parliamentarians. From the extracted tweets, we select those, which refer to newspaper content. This content is then extracted using already existing state-of-the-art methods. Section 4.3 describes this process. Finally, we show some descriptive statistics on the dataset in Section 4.4.

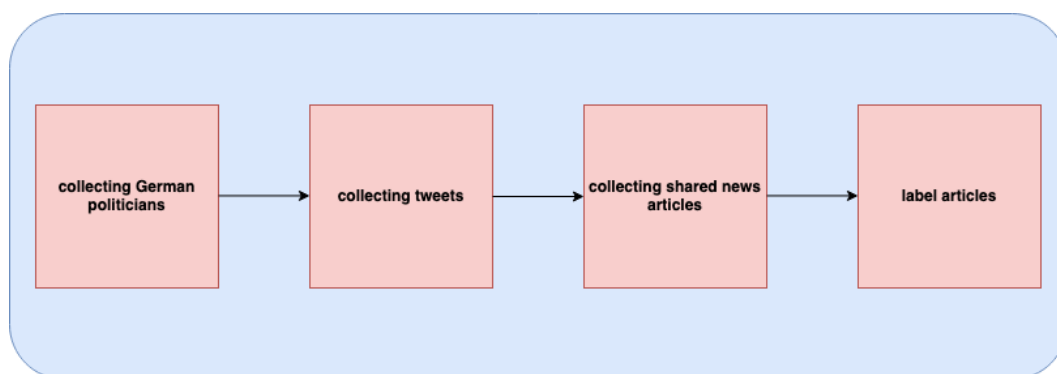
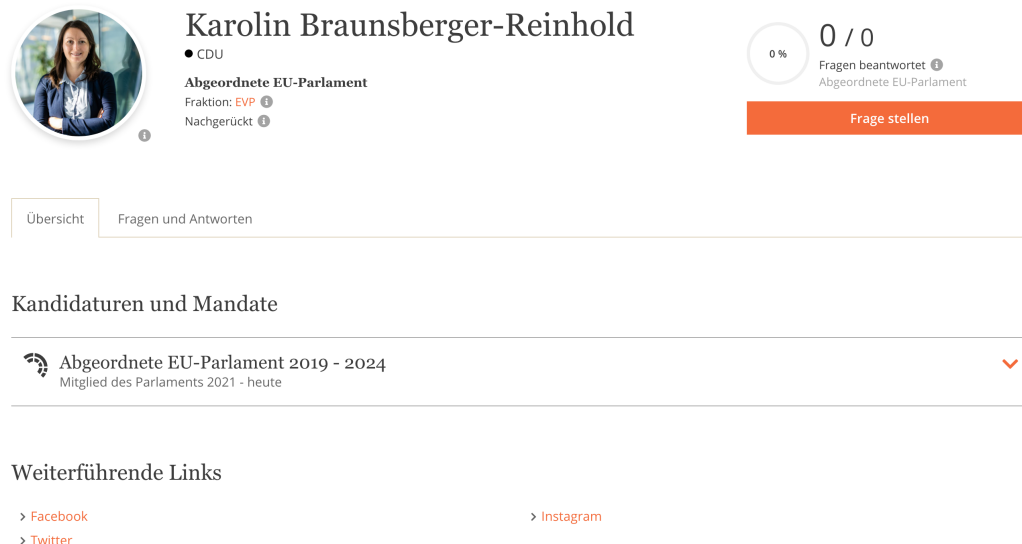


Figure 4.1: Bias Annotation Process by Political Party Affiliation

## 4.1 Collecting Members of Parliaments

Previously in Chapter 1, we introduced the phenomena “echo chamber” and “filter bubble”, which occur when users read and follow the content that aligns with their beliefs. Conversely, it means that users tend to share the content according to the same principle (Freitag et al., 2021; Stefanov et al., 2020; Morgan et al., 2013; Ribeiro et al., 2018; An et al., 2012). Based on this insight, we label the articles according to political leaning of the person who shared the content. Rather than inferring people’s political leanings from assumptions, we identify a specific group of individuals whose political leaning can be clearly inferred: Politicians. The political orientation of the respective politicians is considered common knowledge based on their party affiliation. This assumption needs no further investigation and can be taken as given.

To create a collection of German politicians, we resort to “AbgeordnetenWatch”. “AbgeordnetenWatch” describes themselves as following: “On [abgeordnetenwatch.de](https://abgeordnetenwatch.de) users find a blog, petitions, and short profiles of their representatives in the federal parliament as well as on the state and EU level”.<sup>1</sup> “AbgeordnetenWatch” provides an HTTP API endpoint that delivers the profiles of politicians.



**Karolin Braunsberger-Reinhold**  
 • CDU  
 Abgeordnete EU-Parlament  
 Fraktion:  **EVP**   
 Nachgerückt

0 / 0  
 0 %  
 Fragen beantwortet  
 Abgeordnete EU-Parlament  
 Frage stellen

Übersicht    Fragen und Antworten

Kandidaturen und Mandate

Abgeordnete EU-Parlament 2019 - 2024  
 Mitglied des Parlaments 2021 - heute

Weiterführende Links

> Facebook    > Instagram  
 > Twitter

Figure 4.2: Example profile from AbgeordnetenWatch

The Listing 4.1 illustrates an example profile provided by the specified API endpoint. In addition to the basic information about the respective politician, such as the internal *id*, *name*, *party*, *year of birth* and *sex*, whereby some meta fields do not necessarily have to be filled and are then represented as *null*, the API response also contains the URL, *abgeordnetenwatch\_url*, to the web profile of the respective politician. Since the API unfortunately does not transmit the Twitter ID of the respective politician, we have to extract it from the HTML document of the respective web profile, if available. Figure 4.2 shows an example of a politician’s web profile referenced by *abgeordnetenwatch\_url*. The web profiles on “AbgeordnetenWatch” always include a section called *Related Links*. In this section the politicians provide references to their profiles on social media, e.g., *Facebook*, *Instagram* or *Twitter*.

<sup>1</sup><https://www.abgeordnetenwatch.de/ueber-uns/mehr/international>, accessed: 2022-03-24.

```
1 {
2   "id": "178147",
3   "entity_type": "politician",
4   "label": "Karolin_Braunsberger-Reinhold",
5   "api_url": "https://www.abgeordnetenwatch.de/api/v2/politicians/178147",
6   "abgeordnetenwatch_url":
7     ↪ "https://www.abgeordnetenwatch.de/profile/karolin-...",
8   "first_name": "Karolin",
9   "last_name": "Braunsberger-Reinhold",
10  "birth_name": null,
11  "sex": "f",
12  "year_of_birth": null,
13  "party": "CDU",
14  "party_past": null,
15  "deceased": null,
16  "deceased_date": null,
17  "education": null,
18  "residence": null,
19  "occupation": null,
20  "statistic_questions": null,
21  "statistic_questions_answered": null,
22  "qid_wikidata": null,
23  "field_title": null
24 }
```

Listing 4.1: Example JSON Response from AbgeordnetenWatch API

```
1 {
2   "data": {
3     "created_at": "2015-02-03T23:43:34.000Z",
4     "name": "Karolin_Braunsberger-Reinhold",
5     "username": "kbr_europa",
6     "id": "3015671685",
7     "description": ""
8   }
9 }
```

Listing 4.2: Example JSON Response from Twitter API users endpoint

```
1 {
2   "id": "178147",
3   "entity_type": "politician",
4   "label": "Karolin_Braunsberger-Reinhold",
5   ...
6   "twitter_user_name": "kbr_europa",
7   "twitter_user_id": "3015671685"
8 }
```

Listing 4.3: AbgeordnetenWatch politician profile with twitter profile information

The URL referenced behind the Twitter link looks for example as following: `https://twitter.com/kbr_europa`. The `kbr_europa` part after the last slash — also called *subdirectory* —, represents the Twitter username. We extract these Twitter usernames from "AbgeordnetenWatch" profiles where possible and complete our politician profiles, in order to extract tweets from the Twitter API.

Since you can only extract the tweets from the Twitter API based on the *twitter\_user\_id* of the respective user, we have to extract it based on the *twitter\_user\_name*. Twitter provides an API endpoint<sup>2</sup> for this purpose. Thus, we request the API endpoint for every available *twitter\_user\_name* extracted from "AbgeordnetenWatch". Listing 4.2 shows an exemplary API Response. The information we need is referenced in the key *id*. We append this *id* to our politician profiles with the key *twitter\_user\_id*. Listing 4.3 shows the already existing politician profile, supplemented with two extra keys: *twitter\_user\_name* and *twitter\_user\_id*.

## 4.2 Collecting Tweets

Based on the collection of *twitter\_user\_id* we further collect politician's tweets. For that matter, Twitter provides an HTTP API endpoint<sup>3</sup> enabling users to select additional information about any tweet. Among many possible meta-information, such as media fields, geographic meta-information or even poll fields, we decided to choose a narrow set of meta-information: (1) *entities* and (2) *public metrics*.

**Entities:** Entities are JSON objects providing additional information about references, i.e., *hashtags*, *urls* and *user mentions* used within a Tweet. Every entity itself represents a JSON object consisting at least of start and end index of the entity and the tag itself. Since *mentions* reference another Twitter user, they also include the *id* of that user as additional information. Further, *urls* include the shortened and the expanded url.

**Public metrics:** Public metrics are JSON objects providing additional engagement information, *like\_count*, *quote\_count*, *reply\_count* and *retweet\_count* for a Tweet.

In addition to the described requested meta fields, Twitter User's API endpoint provides information about the time the tweet was created, the tweet ID and the corresponding Tweet text. Listing 4.4 illustrates the exemplary response object from the Tweet API endpoint. For better data handling, we want to simplify and merge the data from "AbgeordnetenWatch" and Twitter and keep only the most necessary information. For this purpose, we merge the personal information of every politician with the information of his or her tweets. Listing 4.5 illustrates the merged object.

<sup>2</sup>[https://api.twitter.com/2/users/by/username/<twitter\\_user\\_name>](https://api.twitter.com/2/users/by/username/<twitter_user_name>)

<sup>3</sup>[https://api.twitter.com/2/users/<twitter\\_user\\_id>/tweets](https://api.twitter.com/2/users/<twitter_user_id>/tweets)

```
1  {
2    "created_at": "2016-07-11T17:21:32.000Z",
3    "entities": {
4      "hashtags": [
5        {
6          "end": 75,
7          "start": 67,
8          "tag": "Studium"
9        },
10       {
11         "end": 85,
12         "start": 80,
13         "tag": "Kind"
14       },
15       {
16         "end": 119,
17         "start": 111,
18         "tag": "Familie"
19       }
20     ],
21     "mentions": [
22       {
23         "end": 17,
24         "id": "276912738",
25         "start": 3,
26         "username": "aufstiegsstip"
27       }
28     ],
29     "urls": [
30       {
31         "display_url": "sueddeutsche.de/bildung/studium-...",
32         "end": 110,
33         "expanded_url": "http://www.sueddeutsche.de/bildung/studium-...",
34         "start": 87,
35         "url": "https://t.co/Fb9bWEnoT2"
36       }
37     ]
38   },
39   "id": "752553424235945985",
40   "public_metrics": {
41     "like_count": 0,
42     "quote_count": 0,
43     "reply_count": 0,
44     "retweet_count": 2
45   },
46   "text": "RT_@aufstiegsstip:_Zwischen_Kinderarzt_und_Vorlesung:_So_
47     ↪  gelingt_das_#Studium_mit_#Kind._https://t.co/Fb9bWEnoT2_#Familie"
```

Listing 4.4: Example JSON Response from Twitter API tweets endpoint

key	description
id	AbgeordnetenWatch internal politician unique id
label	full name of the politician
first_name	first name of the politician
last_name	last name of the politician
sex	gender of the politician
party	party affiliation
twitter_user_name	politician's twitter username
twitter_user_id	politician's unique twitter id
tweet_id	tweet's unique id
created_at	tweets creation time
text	raw tweet text
hashtags	list of hashtags used in the tweet
mentions	list of mentions used in the tweet
urls	list of URLs referenced in the tweet
expanded_urls	list of expanded URLs referenced in the tweet
like_count	count of likes of the tweet
quote_count	count of quoted retweets
reply_count	count of replies
retweet_count	count of retweets

Table 4.1: Description of merged AbgeordnetenWatch and Tweet Information.

```

1 {
2   "id": "178147",
3   "label": "Karolin_Braunsberger-Reinhold",
4   "first_name": "Karolin",
5   "last_name": "Braunsberger-Reinhold",
6   "sex": "f",
7   "party": "CDU",
8   "twitter_user_name": "kbr_europa",
9   "twitter_user_id": "3015671685",
10  "tweet_id": "752553424235945985",
11  "created_at": "2016-07-11T17:21:32.000Z",
12  "text": "RT_@aufstiegsstip:_Zwischen_Kinderarzt_und_Vorlesung:_So_
    ↳ gelingt_das_#Studium_mit_#Kind._https://t.co/Fb9bWEnoT2_#Familie"
13  "hashtags": ["Studium", "Kind", "Familie"],
14  "mentions": ["aufstiegsstip"],
15  "urls": ["https://t.co/Fb9bWEnoT2"],
16  "expanded_urls": [
17    "http://www.sueddeutsche.de/bildung/studium-so-gelingt-..."
18  ],
19  "like_count": 0,
20  "quote_count": 0,
21  "reply_count": 0,
22  "retweet_count": 2
23 }

```

Listing 4.5: Merged AbgeordnetenWatch and Twitter API objects



### 4.3 Collecting News Articles

The collection of tweets, including the extended URL, allows creating a corpus of German-language articles. The extraction of the intended content is a challenging task in computer science and is often referred to as Web Scraping. This task is essential for this thesis, since further analysis as well as training of machine learning models is based on the raw texts as well as the headlines of the linked newspaper articles. Since the development of a web scraper would be a separate and very extensive task, it is not the focus of this work. Therefore, this work draws on already existing system that is able to extract web content. For the purpose of content extraction, in this work we use the tool *trafilatura* (Barbaresi, 2021). This software extracts main text, comments, and metadata of a given news article. According to the evaluation from the paper mentioned above, this tool outperforms other open source solutions. Table 4.2 shows the performance in comparison to other open source tools.

Python Package	Precision	Recall	Accuracy	F-Score	Diff.
justext 2.2.0 (custom)	0.870	0.584	0.749	0.699	6.1x
newspaper3k 0.2.8	0.921	0.574	0.763	0.708	12.9x
boilerpy3 1.0.2	0.851	0.696	0.788	0.766	4.8x
goose3 3.1.9	<b>0.950</b>	0.644	0.806	0.767	18.8x
baseline (text markup)	0.746	0.804	0.766	0.774	1x
dragnet 2.0.4	0.906	0.689	0.810	0.783	3.1x
readability-lxml 0.8.1	0.917	0.716	0.826	0.804	5.9x
news-please 1.5.21	0.924	0.718	0.830	0.808	60x
trafilatura 0.8.2 (fast)	0.925	0.868	0.899	0.896	3.9x
trafilatura 0.8.2	0.934	<b>0.890</b>	<b>0.914</b>	<b>0.912</b>	8.4x

Table 4.2: Performance of *trafilatura* in comparison to other web scraping tools.<sup>4</sup>

The procedure of extracting articles looks as follows: (1) We iterate over all politician merged with the corresponding Tweet object, Listing 4.5; (2) Extract the content of every expanded URL listed in *expanded\_urls*; (3) Store the content as JSON object with corresponding *tweet\_id* as filename. Listing 4.6 shows an example JSON object with main content extracted with *trafilatura*. The most important fields in the object are: *title*, *hostname*, *raw-text*, and *excerpt* — the short description of the article, often mentioned on top of every article. Additionally, the object contains some other useful information like *categories*, *tags* and *comments*. The extraction also include some meta information, e.g., publishing *date* of the article, *id* and an automatically generated unique fingerprint of the document.

We merge the extracted information from articles with corresponding politician’s tweets objects listed in Listing 4.5. In the end, we have a table, where every row represents information about the tweet and the author of the tweet as well as the extracted main content from the referenced URL in the tweet.

<sup>4</sup><https://trafilatura.readthedocs.io/en/latest/index.html#evaluation-and-alternatives>, accessed: 2022-03-31.

```

1 {
2   "title": "So_gelingt_das_Studium_mit_Kind",
3   "author": "",
4   "hostname": "sueddeutsche.de",
5   "date": "2016-06-28",
6   "categories": "",
7   "tags": "",
8   "fingerprint": "EFr6xinI7gtjqBbXj5bGuoSCBa4=",
9   "id": null,
10  "license": null,
11  "raw-text": "Nach_gerade_mal_drei,_vier_Stunden_Schlaf_um_sechs...",
12  "source": "https://www.sueddeutsche.de/bildung/studium-so-gelingt-...",
13  "source-hostname": "sueddeutsche.de",
14  "excerpt": "Laut_20._Sozialerhebung_des_Deutschen_Studentenwerks...",
15  "comments": ""
16 }

```

Listing 4.6: Example article extracted with trafilatura

## 4.4 Dataset Statistics

The created raw dataset consists of 44,167 unique tweets with corresponding unique news articles. Due to the fact, that every web scraping tool, including *trafilatura* is not perfect and has some error rate in extracting main content, we filter the dataset. The filtering process consists of two steps: (1) remove rows where *title*, *excerpt* or *raw\_text* could not be extracted; (2) remove rows where *excerpt* is too short. While the first step is simple, we need to take a closer look at the length distribution of the excerpts in the second step.

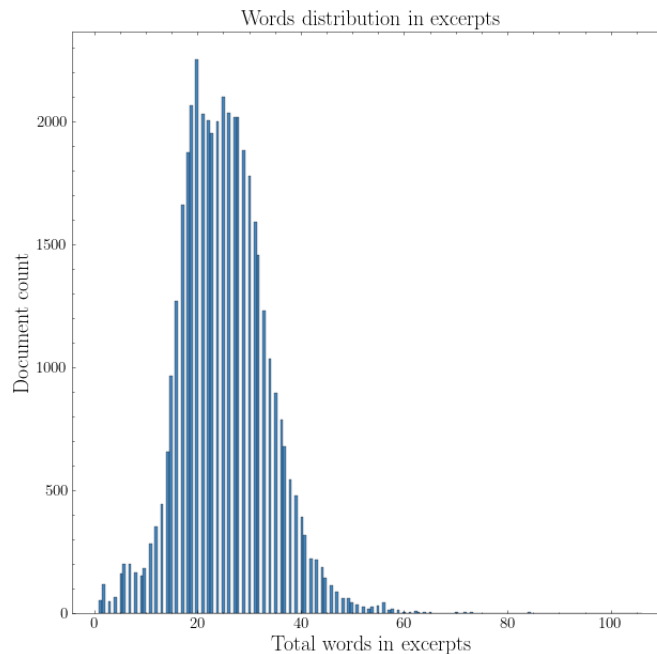


Figure 4.3: Words distribution in articles excerpts.

Figure 4.3 shows the length distribution of extracted excerpts. The calculation of quantiles shows that 5% of the excerpts consists of less than 13 words or 98 characters. Random examination of the subset of summaries revealed that they were either mistakenly not extracted in their entirety, or that they were actually so short and had very little information content. Therefore, we decided to filter them out. In comparison, Twitter used to only allow tweets of less than 120 characters and was later raised to 240 characters.

Party	Politicians count	Tweets count	Mean Tweets/Politician
AfD	111	7354	<b>66</b>
DIE GRÜNEN	<b>327</b>	<b>10438</b>	32
DIE LINKE	173	5560	32
FDP	204	7158	35
SPD	296	6246	21
UNION	197	4819	25
<b>Total</b>	1308	41575	

Table 4.3: Number of MPs and tweets per party.

Table 4.3 shows further details of constructed dataset: the total count of politicians and tweets per party. Looking at the table, we can see that most of the MPs in the dataset belong to the party "DIE GRÜNEN", followed by the members of the "SPD". With 10,438 tweets, "DIE GRÜNEN" is also clearly ahead of the competition in terms of the number of tweets. However, the ratio of the number of tweets to the number of MPs looks different: while the average number of tweets per party ranges from a minimum of 21 ("SPD") to a maximum of 32 ("DIE GRÜNEN", "DIE LINKE"), the "AfD" tweets significantly more with around 66 tweets per MP.

Furthermore, we would like to look at the more detailed breakdown of shared articles per party per publisher. For that purpose, we illustrate the insights in Table 4.4: every column represents the tweets of the corresponding party and sums up to total number of tweets of the party. Every row corresponds to each publisher. Based on the table, you can see which party shares which medium the most on Twitter. The most shared publishers on Twitter are **WELT**, **SPIEGEL** and **SZ**.

	AFD	DIE GRÜNEN	DIE LINKE	FDP	SPD	UNION
BILD	997	155	108	618	244	721
DFUNK <sup>5</sup>	69	312	177	115	203	144
FAZ	1109	908	325	1478	645	850
FOCUS	955	176	141	425	137	304
N-TV	538	281	332	474	237	320
SPIEGEL	552	1939	<b>1176</b>	813	<b>1549</b>	459
STERN	57	108	76	76	107	54
SZ	274	<b>2254</b>	708	572	1090	357
TAGESSCHAU	122	421	348	211	269	142
TAZ	80	1553	868	96	319	60
WELT	<b>2259</b>	687	620	<b>1607</b>	585	<b>1062</b>
ZEIT	291	1472	570	571	751	274
ZDF	51	172	111	102	110	72
<b>Total</b>	7354	10438	5560	7158	6246	4819

Table 4.4: Absolute number of shared articles per party, per publisher in alphabetical order.

---

<sup>5</sup>Deutschlandfunk

This chapter describes the standard machine learning process in context of *natural language processing* (NLP) tasks illustrated in Figure 5.1.

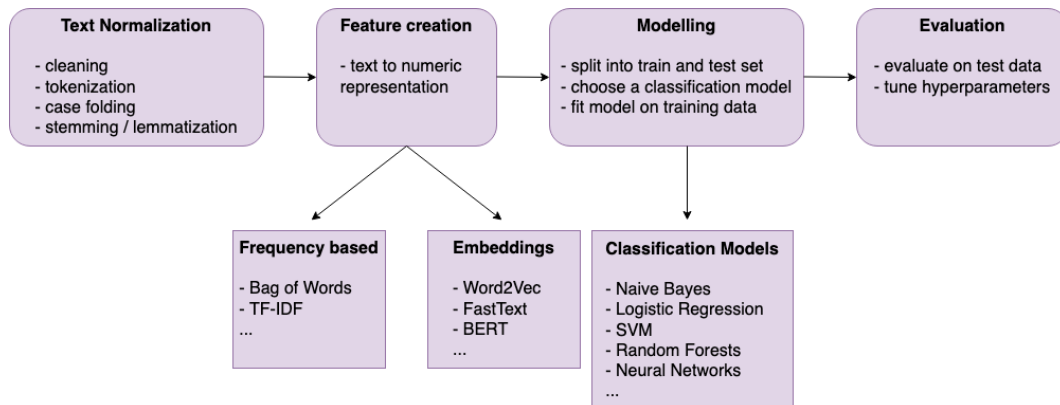


Figure 5.1: NLP machine learning workflow

In the first step, we define the classification problem and the research questions regarding the collected data. Within the scope of experiment settings, we then describe the process of text normalization, including cleaning, tokenization, case folding and lemmatizing. Further, we describe the process of data-splitting, including the concepts of random over- and under-sampling techniques, for each research question, respectively. Moreover, we present used techniques to translate the text into a numerical representation, frequency based TD-IDF and the embeddings, depending on model selection. Moreover, we then describe the experiment settings and the algorithms we train to approach the defined research questions. We train three standard machine learning models, *Logistic Regression*, *Linear Support Vector* and *Naive Bayes*, on

frequency-based text representation. Additionally, we train a more sophisticated state-of-the-art model, a Bidirectional Encoder Representations from Transformers, *BERT* (Devlin et al., 2019), a transformer-based machine learning technique with *embeddings* as features. The subsequent evaluation and the interpretation of the results is provided in Chapter 6.

## 5.1 Research Questions

Considering the dataset from Chapter 4, we define a classification problem as follows: Given a textual content, e.g., *title*, *excerpt* or *raw text* of an article, estimate the political orientation of that article. We define the political orientation of articles as labels  $y$  for  $y \in \{\text{AfD}, \text{DIE GRÜNEN}, \text{DIE LINKE}, \text{FDP}, \text{SPD}, \text{UNION}\}$ , according to the political affiliation of the politician who shared the article on Twitter. We formulate the following research questions based on the above defined classification task:

- Q1.** How effective are standard classification approaches with frequency-based features and the standard neural approach in detecting bias from the right-wing political fringe, when using the articles shared by all other parties as negative samples?
- Q2.** How effective are the same standard approaches in detecting bias from right-wing political fringe, when training against the articles shared by other parties separately? As in the first research question, we label articles shared by *AfD* politicians as biased and the others as unbiased.
- Q3.** How do the same approaches perform when applying to the detection of the left-wing party *DIE LINKE*, i.e.: a) when using the articles shared by all other parties as negative samples, analogous to **Q1**; b) when training articles of *DIE LINKE* separately against articles of every other party, analogous to **Q2**.

## 5.2 Experiment Settings

### Text Normalization

Like for almost any approach of natural language processing of a text, e.g., text classification like in our case, we have to perform a **text normalization** in the first step. At the end of the text normalization step, the texts are mapped into a vector space model representation. In the vector space model, each token, e.g., word, represents one dimension. The respective document is represented as a vector in the multidimensional space. The number of unique words corresponds to the number of dimensions. Various text normalization steps are used to reduce the dimensionality, which often leads to model's performance improvement. In the following, we describe the normalization steps we apply to the excerpts of German news articles.

**Cleaning** Since the excerpts are automatically extracted texts from HTML documents, we have to make sure that unnecessary characters are not accidentally extracted. Therefore, we remove all possible HTML tags (e.g., `</br>`) from the texts, keep only ASCII characters, remove single letter chars and convert all white spaces (e.g., tabs) to single white spaces. For frequency based feature modelling, we also remove any

punctuation marks, as they are meaningless in that case. We use the *NLTK*<sup>1</sup> library to remove stop words from excerpts.

**Tokenization** Tokenization is the process of dividing the text into smaller segments, called *tokens*, e.g., *words*, *characters*, or *n*-grams. The need for tokenization lies in the fact that tokens represent the basic building blocks of every language, and the common methods of raw text processing is based on the token level. For standard featured-based approaches, we tokenize on word level.

**Case Folding** Case folding represents a method of word normalization. For some tasks, e.g., speech recognition and information retrieval, and some languages mapping every word to lower case can be very useful for generalization. However, since in our case we are processing German-language texts where there is a distinction between upper and lower case, we keep the spelling and intentionally omit this step.

original word	stem	lemma
Entscheidung	Entscheid	Entscheidung
Netzneutralität	Neutzneutralitat	Netzneutralität
Umsetzung	Umsetz	Umsetzung

Table 5.1: Comparison between *stemming* and *lemmatization*.

**Lemmatization** Two other methods are also part of the word normalization: *stemming* and *lemmatization*. During *stemming*, the words are truncated to their base/root form. A major disadvantage of stemming is that it sometimes truncates words, losing the original meaning or failing to produce a proper word in a given language. In contrast, *lemmatization* cuts the words in such a way that the original meaning is retained. In lemmatization, one uses pre-defined dictionaries that store the context of the words, and checks the word as it is truncated. The examples in Table 5.1 illustrate the advantages of *lemmatization* over *stemming*. For the above reasons, we prefer to lemmatize rather than stem the text for our experiments.

		Training		Test	
		Bias	Neutral	Bias	Neutral
<b>Q1</b>	AfD-vs-REST	5515	25666	1839	8555
	AfD-vs-DIE GRÜNEN	5515	7829	1839	2609
	AfD-vs-DIE LINKE	5515	4170	1839	1390
<b>Q2</b>	AfD-vs-FDP	5515	5369	1839	1789
	AfD-vs-SPD	5515	4685	1839	1561
	AfD-vs-UNION	5515	3614	1839	1205

Table 5.2: Bias distribution of excerpts in Q1 and Q2.

<sup>1</sup>Natural Language Toolkit <https://www.nltk.org/>, accessed: 2022-05-04

		Training		Test	
		Bias	Neutral	Bias	Neutral
<b>Q3(a)</b>	DIE LINKE-vs-REST	4170	27011	1390	9004
	DIE LINKE-vs-DIE GRÜNEN	4170	7829	1390	2609
	DIE LINKE-vs-AfD	4170	5515	1390	1839
<b>Q3(b)</b>	DIE LINKE-vs-FDP	4170	5369	1390	1789
	DIE LINKE-vs-SPD	4170	4685	1390	1561
	DIE LINKE-vs-UNION	4170	3614	1390	1205

Table 5.3: Bias distribution of excerpts in Q3.

To study the defined research questions, we split the raw data collected in Chapter 4 in two different subsets. In the following, we describe the dataset for each research question, respectively.

**Data splits for Q1.** To train the models to differentiate between shared articles from right-wing politicians and the others, we partition the data in two classes: articles that were shared from *AfD* are labeled as biased, i.e., class 1 and articles that were shared from any of other parties as unbiased with class 0. We then split the entire dataset in train (75%) and test (25%) split, having 5.515 excerpts as biased and 25.666 as neutral in the training set, and 1.839 biased and 8.555 neutral excerpts in the test set.

**Data splits for Q2.** For the second research question, we build five subsets of the entire dataset: every subset contains the articles shared by the right-wing *AfD* or by exactly one of the other parties, e.g.,  $s_1 = \{a_i \in AfD | a_i \in SPD\}$ ,  $s_2 = \{a_i \in AfD | a_i \in FDP\}$ , etc. We label the examples the same way as in the first research question: articles shared by *AfD* as *biased*, articles shared by any of the other party as *unbiased*.

**Data splits for Q3.** In the third research question, we differentiate two cases:

- a) Analogous to the first research question, we partition the data in two classes: articles that were shared from *DIE LINKE* are labeled as biased, i.e., class 1 and articles that were shared from any of other parties as unbiased with class 0. We then split the entire dataset in train (75%) and test (25%) split.
- b) Analogous to the second research question, we build five subsets of the entire dataset: every subset contains the articles shared by the left-wing *DIE LINKE* or by exactly one of the other parties, e.g.,  $s_1 = \{a_i \in DIE LINKE | a_i \in SPD\}$ ,  $s_2 = \{a_i \in DIE LINKE | a_i \in FDP\}$ , etc. We label the examples the same way as in the Q3a scenario: articles shared by *DIE LINKE* as *biased*, articles shared by any of the other party as *unbiased*.

Table 5.2 shows the distribution of bias and non-biased excerpts in train and test sets for the first and second research questions. Table 5.3 shows the class distribution in training and test sets for the third research question, for both scenario a and scenario b. Due to the high imbalance in data distribution for **Q1** and **Q3a**, we experiment



with under- and over-sampling strategies of the data and compare the results. In the following, we describe the concepts of both strategies:

**Random Under-sampling:** Refers to a technique that samples records randomly from the majority class. The process terminates as soon as the records from both class are balanced. For under-sampling, we use the *RandomUnderSampler* class from the *imblearn*-library<sup>2</sup>.

**Random Over-sampling:** Refers to a technique that duplicates sampled from the minority class. The process terminates as soon as the balance is reached regarding the majority class. For over-sampling, we do duplicate the samples from the minority class using the *RandomOverSampler* class from *imblearn*-library<sup>3</sup>.

## Feature Creation and Modelling

Before applying any classification model on textual data, text have to be converted into its numerical representation. The simplest form to represent a text into a numerical way is called *Bag of Words*. For this purpose, we first create a list of the vocabulary used in the entire document collection, assigning every word an index. Further, every sentence is represented as a vector of length  $N$ , size of the vocabulary, indicating the presence of each word from dictionary with 1 at the corresponding index. All words that do not appear in the sentence are represented as 0. However, this easy-to-implement method has many disadvantages compared to other options. The resulting matrix of documents, where every row is a document and every column is a corresponding word from the vocabulary, is very sparse, containing less information. Another problem with this representation is that it contains no information about the grammar of the sentences, ordering of the words, nor any meaning of the importance of each word for each document and the entire corpus. Having provided an intuition for how text can be translated into a numerical representation, we would like to discuss the approaches we use in this work: **term frequency-inverse document frequency (TF-IDF)** and **Embeddings**.

**TF-IDF** TF-IDF is a measure from information retrieval, that estimates the relevance of each token, e.g., word or in general  $n$ -gram in a single document within the entire document collection. TF-IDF consists of two concepts: *term frequency (TF)* and *inverse document frequency (IDF)*. For the calculation, we first define a set  $E$  containing all excerpts  $e_1, \dots, e_n$ , with  $e_n$  indicating the last excerpt in the document collection:

$$E = \{e_1 \dots e_n\} \quad (5.1)$$

Every excerpt consists of each individual collections of tokens  $t$ , where the token  $t$  can represent each single word (unigram), or in general a combination of subsequent words as  $n$ -grams:

$$e_i = \{t_1 \dots t_m\} \quad (5.2)$$

<sup>2</sup>[https://imbalanced-learn.org/stable/references/generated/imblearn.under\\_sampling.RandomUnderSampler.html](https://imbalanced-learn.org/stable/references/generated/imblearn.under_sampling.RandomUnderSampler.html), accessed: 2022-05-04

<sup>3</sup>[https://imbalanced-learn.org/stable/references/generated/imblearn.over\\_sampling.RandomOverSampler.html](https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.RandomOverSampler.html), accessed: 2022-05-04

, where  $t_m$  indicates the last token in each excerpt respectively. The TF part represents the term frequency of each token for a given document. The IDF for each token  $t_j$  is calculated by

$$IDF_{t_j} = \log \frac{N}{df_{t_j}} \quad (5.3)$$

, where  $N$  represents the total number of all documents in a collection, and  $df_{t_j}$  the total number of documents, that contains the term  $t_j$ . The final TF-IDF score for each token in a document is calculated by multiplying the *term frequency* of each token in a given document  $e_i$  and the IDF score and its *inverse document frequency*.

$$TF-IDF_{t_i} = TF_{e_i}(t_j) \cdot IDF_{t_j} \quad (5.4)$$

Unlike weighing the *n-grams*, e.g., uni-grams, bi-grams, etc., just by their frequency in a document, *TF-IDF* controls the importance of an *n-gram* by calculating its frequency of occurrence in a document and normalizing by the frequency of occurrence in the whole document collection. In that way, the method assigns more weight, or importance, to words that occur many times in a single document, but less in the whole document collection, meaning that this word might be relevant for the specific document. From the other perspective, words that often occur in many documents, become less important.

As standard feature-based approaches for answering **Q1.** and **Q2.**, we employ a *Linear Support Vector*, *Logistic Regression* and a *Multinomial Naive Bayes* classifiers based on **TF-IDF** vectors of word *n-grams* with  $n \in \{1, 2, 3\}$  as a baseline. Due to the limited number of data records for each class, we omit to create the validation data set and instead perform a *StratifiedKFold* technique with  $k = 5$ , to validate the classifiers during training.

**Embeddings** Instead of representing words by a single number as in TF-IDF, the basic idea of *embeddings* is to use a vector representation (list of numbers) for each word, which in a certain way represent the semantic, i.e., recognizing whether the words are similar or opposites, and syntactic relations, i.e., recognizing that the words “have” and “has” have the same relation as “be” and “is”. The first two methods that implemented the idea of *word embeddings* to create pre-trained representations of the distribution of words were *Word2Vec* (Mikolov et al., 2013) and *GloVe* (Pennington et al., 2014). A well-known drawback of the above approaches is that, unlike real languages, word embeddings are always the same regardless of context. Thus, this thesis make use of *contextualized embeddings* generated by *Bidirectional Encoder Representations from Transformers*, **BERT**, (Devlin et al., 2019) and typically refer to the output of the final layer of a stacked Transformer (Vaswani et al., 2017) architecture. The conventional workflow consists of two separate stages: (1) pre-training using two self-supervised tasks, and (2) fine-tuning for downstream applications. In the pre-training stage, **BERT** uses two self-supervised tasks: *masked language modeling*, where the goal is to make a prediction from randomly masked input tokens, and *next sentence prediction*, determining if two sentences are adjacent to each other. In the fine-tuning stage for downstream tasks, e.g., *classification*, the standard approach is to add one or more fully connected layers on top of the final encoder layer. This thesis employs a pre-trained German cased *BERT* model using *contextualized BERT word embeddings* as features. The German cased *BERT* model<sup>4</sup> is trained on a huge amount of data: Wiki, OpenLegalData and

<sup>4</sup><https://huggingface.co/bert-base-german-cased>, accessed: 2022-05-04

News (~12 GB). We divide the dataset in 80% training, 10% validation and 10% test sets. We fine-tune the model using the training data set for each research question and evaluate on the test dataset, respectively. Since the BERT model accepts maximum 512 tokens as sequence length, we use the excerpts of the articles instead of the entire news articles. Since providing a deeper understanding of *BERT*, stacked *Transformer Architecture* with self-attention and the pre-training stage is beyond the scope of this paper, we refer to the original papers Devlin et al. (2019) and Vaswani et al. (2017).



The present chapter describes the evaluation of the approaches described in Chapter 5. Firstly, we present the evaluation metrics that we use to evaluate the trained models. We then describe the results of the first research question and explain the performance difference between *n-gram* and the *embedding* models. Furthermore, we illustrate the results of the second research question and emphasize the performance difference in comparison to the first research question. With the results of the experiments regarding the third research question, we show that the approaches work for the detection of the articles with left political leaning too. In the end, we briefly give an interpretation of the results.

### Evaluation Metrics

In the following, we present the metrics we use to evaluate the results of experiments regarding the defined research questions.

**Precision:** Precision, or the positive predictive value, describes a measure of a classifiers' exactness, in other words: how many samples classified as positive are true positive.

$$Precision = \frac{tp}{tp + fp} \quad (6.1)$$

**Recall:** Recall or the true positive rate, sensitivity, describes a measure of a classifiers' completeness, in other words: from all positive samples, how many does the classifier recognized correctly as such.

$$Recall = \frac{tp}{tp + fn} \quad (6.2)$$

**F<sub>1</sub>-Score:** F<sub>1</sub>-Score combines precision and recall and represents the harmonic mean of both metrics. Since we want both high precision and high recall, the F-Score seems

to be a good metric that maps them both.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6.3)$$

## Q1. Results

Sampling	Feature	Classifier	Prec.	Recall	F1
Original Distribution	n-gram(1-3)	Logistic Regression	0.44	0.15	0.22
	n-gram(1-3)	Linear Support Vector	0.31	0.30	0.30
	n-gram(1-3)	Multinomial NB	0.33	0.30	0.31
	Embeddings	BERT	0.61	0.32	<b>0.42</b>
Over-Sampling	n-gram(1-3)	Logistic Regression	0.29	0.55	0.39
	n-gram(1-3)	Linear Support Vector	0.29	0.55	0.39
	n-gram(1-3)	Multinomial NB	0.31	0.56	0.40
	Embeddings	BERT	0.63	0.56	<b>0.59</b>
Under-Sampling	n-gram(1-3)	Logistic Regression	0.30	0.52	0.38
	n-gram(1-3)	Linear Support Vector	0.29	0.55	0.38
	n-gram(1-3)	Multinomial NB	0.27	0.56	0.36
	Embeddings	BERT	0.61	0.57	<b>0.59</b>

Table 6.1: Performance comparison of classifiers regarding the first research question. *Multinomial NB* corresponds to the Multinomial Naive Bayes Classifier. We perform *TF-IDF* on the extracted *n-grams(1-3)*. BERT is fine-tuned with a learning rate of  $2e^{-6}$  and shows no improvement in validation loss and validation accuracy after the 3rd epoch.

**Research Question:** How effective are standard classification approaches with frequency-based features and the standard neural approach in detecting bias from the right-wing political fringe, when using the articles shared by all other parties as negative samples?

We analyze the first research question using three possible data distribution options: (1) original unbalanced data distribution across the defined classes; (2) random over-sampling data, i.e., create randomly duplication of data records from the minority class; (3) random under-sampling, i.e., randomly removing data records from the majority class. The Table 6.1 illustrates the results regarding the first research question. In the unbalanced scenario, all models trained on TF-IDF of *n-grams(1-3)* suffer from class distribution in favor of the negative class (the unbiased majority class), in other words: the models simply tend to predict the majority class. This causes the recall of the positive class to be so low: the samples that are positive are predicted to be negative. Since we optimize the models towards the F1-Score, it results in similar performance in precision and recall. The best performance in the unbalanced scenario is shown by the BERT model with a precision of 0.61, the model also manages to correctly detect over 40% of the positive samples (recall=0.42). An obvious explanation in the performance difference of the models is essentially what the models were trained on: the TF-IDF feature space might already have a bias towards the majority class that propagates. BERT, on the other hand, was trained on a much larger and more heterogeneous dataset and was only fine-tuned on our dataset.

Party pairs	Feature	Classifier	Prec.	Recall	F1
AfD-DIE GRÜNEN	n-gram(1-3)	Logistic Regression	0.72	0.64	0.68
	n-gram(1-3)	Linear Support Vector	0.72	0.64	0.68
	n-gram(1-3)	Multinomial NB	0.72	0.73	0.72
	Embeddings	BERT	0.72	0.79	<b>0.75</b>
AfD-DIE LINKE	n-gram(1-3)	Logistic Regression	0.73	0.71	0.72
	n-gram(1-3)	Linear Support Vector	0.73	0.71	0.72
	n-gram(1-3)	Multinomial NB	0.72	0.75	0.74
	Embeddings	BERT	0.74	0.76	<b>0.75</b>
AfD-FDP	n-gram(1-3)	Logistic Regression	0.68	0.63	0.65
	n-gram(1-3)	Linear Support Vector	0.69	0.63	0.66
	n-gram(1-3)	Multinomial NB	0.67	0.66	0.67
	Embeddings	BERT	0.71	0.66	<b>0.69</b>
AfD-SPD	n-gram(1-3)	Logistic Regression	0.68	0.66	0.67
	n-gram(1-3)	Linear Support Vector	0.69	0.65	0.67
	n-gram(1-3)	Multinomial NB	0.66	0.67	0.67
	Embeddings	BERT	0.70	0.70	<b>0.70</b>
AfD-UNION	n-gram(1-3)	Logistic Regression	0.62	0.66	0.64
	n-gram(1-3)	Linear Support Vector	0.62	0.66	0.64
	n-gram(1-3)	Multinomial NB	0.62	0.65	0.63
	Embeddings	BERT	0.70	0.67	<b>0.68</b>

Table 6.2: Performance comparison of classifiers regarding the second research question. We perform *TF-IDF* on the extracted *n-grams(1-3)*. BERT is fine-tuned with a learning rate of  $2e^{-6}$  and shows no improvement in validation loss and validation accuracy after the 3rd epoch.

Both balancing methods show similar tendency regarding the results for models trained on TF-IDF of *n-grams*. With both balancing methods, the precision decreases, but the recall increases. Therefore, it can be concluded that the algorithms in both cases tend to predict samples as positive, i.e., biased. This increases the rate of false positives, which explains the lower precision, but decreases the rate of false negatives, which in turn justifies the higher recall. In both balancing scenarios, however, BERT shows the best performance regarding the defined metrics. The BERT manages to optimize recall, but not at the expense of precision. BERT achieves the best performance with oversampling with precision of 0.63, recall of 0.56 and F1-score of 0.59.

## Q2. Results

**Research Question:** How effective are the same standard approaches in detecting bias from right-wing political fringe, when training against the articles shared by other parties separately? As in the first research question, we label articles shared by *AfD* politicians as biased and the others as unbiased.

We obtain more promising results when looking at the experiments regarding the second research question. Here we do not fully balance the data because the inequality in the class distribution is minimal. The classifiers manage better to distinguish *AfD* from other parties in a direct comparison than when the other parties are grouped into a set, like in the first experiment. Table 6.2 illustrates the detailed results of the ex-

periments regarding the second research question. In comparison to the results of the first research question, we observe a strong performance boost. Regarding the models trained on *n-grams*, the precision lies between min 0.62 (AfD-UNION) and max of 0.73 (AfD-DIE LINKE). The lowest recall of 0.63 is achieved with the logistic regression on the AfD-FDP subset, while the highest recall is 0.75 when comparing AfD and DIE LINKE. Again, however, as with the first research question, BERT achieves the best performance regardless of which subset of the data we consider. The F1-scores of BERT varies between 0.68 (AfD-UNION) and 0.75 (AfD-DIE LINKE and AfD-DIE GRÜNEN). Based on the performance of BERT, we can see that the model manages to minimize both false positives and false negatives, thus increasing both metrics, precision and recall to a similar degree. Moreover, the performance of the *BERT* model shows some direct dependence on the political proximity between compared parties: the model performs better in classifying the articles shared by the right-wing *AfD* when regarding only the articles of *AfD* and of the German green party *DIE GRÜNEN* or of the left party *DIE LINKE*, and shows worse performance when classifying with the more conservative *UNION*.

### Q3. Results

**Research Question:** How do the same approaches perform when applying to the detection of the left-wing party *DIE LINKE*, i.e.:

- How effective are standard classification approaches with frequency-based features and the standard neural approach in detecting bias from the left-wing political fringe, when using the articles shared by all other parties as negative samples?
- How effective are the same standard approaches in detecting bias from left-wing political fringe, when training against the articles shared by other parties separately? As in **Q3a**, we label articles shared by politicians of *DIE LINKE* as biased and the others as unbiased.

Sampling	Feature	Classifier	Prec.	Recall	F1
Original Distribution	n-gram(1-3)	Logistic Regression	0.22	0.54	0.31
	n-gram(1-3)	Linear Support Vector	0.25	0.52	0.34
	n-gram(1-3)	Multinomial NB	0.27	0.39	0.32
	Embeddings	BERT	0.58	0.28	<b>0.38</b>
Over-Sampling	n-gram(1-3)	Logistic Regression	0.25	0.49	0.33
	n-gram(1-3)	Linear Support Vector	0.25	0.50	0.33
	n-gram(1-3)	Multinomial NB	0.24	0.53	0.33
	Embeddings	BERT	0.55	0.57	<b>0.56</b>
Under-Sampling	n-gram(1-3)	Logistic Regression	0.21	0.55	0.30
	n-gram(1-3)	Linear Support Vector	0.21	0.55	0.30
	n-gram(1-3)	Multinomial NB	0.21	0.58	0.31
	Embeddings	BERT	0.54	0.59	<b>0.56</b>

Table 6.3: Performance comparison of classifiers regarding **Q3a**. *Multinomial NB* corresponds to the Multinomial Naive Bayes Classifier. We perform *TF-IDF* on the extracted *n-grams(1-3)*. BERT is fine-tuned with a learning rate of  $2e^{-6}$  and shows no improvement in validation loss and validation accuracy after the 3rd epoch.



Party pairs	Feature	Classifier	Prec.	Recall	F1
DIE LINKE-DIE GRÜNEN	n-gram(1-3)	LR	0.45	0.74	0.56
	n-gram(1-3)	Linear SV	0.48	0.64	0.55
	n-gram(1-3)	Multinomial NB	0.49	0.60	0.54
	Embeddings	BERT	0.60	0.59	<b>0.59</b>
DIE LINKE-AfD	n-gram(1-3)	LR	0.63	0.68	0.65
	n-gram(1-3)	Linear SV	0.62	0.70	0.66
	n-gram(1-3)	Multinomial NB	0.61	0.66	0.63
	Embeddings	BERT	0.71	0.73	<b>0.72</b>
DIE LINKE-FDP	n-gram(1-3)	LR	0.61	0.65	0.63
	n-gram(1-3)	Linear SV	0.62	0.64	0.63
	n-gram(1-3)	Multinomial NB	0.63	0.61	0.62
	Embeddings	BERT	0.65	0.62	<b>0.64</b>
DIE LINKE-SPD	n-gram(1-3)	LR	0.60	0.58	0.59
	n-gram(1-3)	Linear SV	0.56	0.62	0.59
	n-gram(1-3)	Multinomial NB	0.63	0.58	0.60
	Embeddings	BERT	0.63	0.59	<b>0.61</b>
DIE LINKE-UNION	n-gram(1-3)	LR	0.65	0.71	0.67
	n-gram(1-3)	Linear SV	0.62	0.69	0.65
	n-gram(1-3)	Multinomial NB	0.69	0.70	0.69
	Embeddings	BERT	0.68	0.73	<b>0.70</b>

Table 6.4: Performance comparison of classifiers regarding **Q3b**. We perform *TF-IDF* on the extracted *n-grams(1-3)*. BERT is fine-tuned with a learning rate of  $2e^{-6}$  and shows no improvement in validation loss and validation accuracy after the 3rd epoch. *LR* corresponds to the Logistic Regression Classifier, *Linear SV* to the Linear Support Vector Classifier and *Multinomial NB* to the Multinomial Naive Bayes Classifier.

Table 6.3 shows the results regarding **Q3a**. Analogous to the experiments regarding the first research question, we evaluate the performances of the classifiers on the whole, random over-sampled and random under-sampled datasets. In comparison to the results of the experiments regarding **Q1**, the classifiers trained on TF-IDF of *n-grams(1-3)* tend to better overcome the imbalanced class distribution due to the higher recall, i.e., the classifiers do not just predict the majority class. Nevertheless, the precision is still very low, i.e., the classifiers fail to precisely detect the articles from the left-wing party *DIE LINKE* when trained on the whole dataset, and misclassify the records, resulting in high rate of *false positives*. The best, but still low, performance regarding the *F1-score* achieves the *BERT* model, with a F1-Score of 0.38, but still having a high *false negative* score, meaning that the *BERT* model still struggles with the high imbalanced class distribution. In both balancing scenarios, however, *BERT* model shows the best performance improvement regarding the defined metrics. The *BERT* model manages to optimize recall, but not at the expense of precision. *BERT* achieves similar performance in random over-sampling and under-sampling, with a F1-score of 0.56.

Similar to **Q2**, the results of **Q3a** illustrated in Table 6.4 are more meaningful. Regarding the models trained on *n-grams*, the precision lies between min 0.49 (DIE LINKE-DIE GRÜNEN) and max of 0.69 (DIE LINKE-UNION). The lowest recall of 0.58 is achieved with the logistic regression on the DIE LINKE-SPD subset, while the highest recall is 0.74 when comparing DIE LINKE and DIE GRÜNEN. Again, however, similar

to **Q1**, **Q2** and **Q3a**, BERT achieves the best performance regardless of which subset of the data we consider. The F1-scores of BERT varies between 0.59 (DIE LINKE-DIE GRÜNEN) and 0.75 (DIE LINKE-AfD). Again, the results show a similar dependency as in Q2: *BERT* model performs better in classifying the articles shared by the left-wing *DIE LINKE* when regarding only the articles of *DIE LINKE* and of the right-wing party *AfD*, and shows worse performance when classifying with the green party (DIE GRÜNEN) or with the left party (DIE LINKE).

**Summary** The findings of the evaluation are multifaceted. Regarding **Q1** and **Q3a**, we can state that the performance of the models suffers strongly from the disbalance of the class distribution. However, the BERT model performs best in all variants of the balancing. This suggests that the model can identify certain meaningful features to distinguish the two classes, even if they are not yet sufficient for above-average performance. The fact that we lump the texts disseminated by all parties, in **Q1** except the *AfD* and except *DIE LINKE* in **Q3a**, together, may also be a reason for this performance, as the underlying concepts of the texts are mixed up, making it difficult to distinguish them from the texts shared by the *AfD* or *DIE LINKE*, respectively. This assumption is in a way confirmed by the results of the second approach in **Q2** and **Q3b**. We see that the models perform significantly better in the direct comparison to every party separately. With almost comparable and remarkable performance in **Q2**, the models ( $0.68 \leq BERT_{F1} \leq 0.75$ ) manage to distinguish the concepts of texts shared by *AfD* from them of the other parties. The results of the approach regarding **Q3b** show a similar tendency with  $0.59 \leq BERT_{F1} \leq 0.72$ . Moreover, the results show that the performance of the models in **Q2** and **Q3b** seems to be directly related to the proximity of the respective parties being compared: in **Q2** we observe, that the *BERT* model performs better in correctly classifying the articles shared by *AfD* in comparison to *DIE LINKE* or *DIE GRÜNEN* and worse in comparison to the more conservative *UNION*. Similar to that, in **Q3b** the *BERT* model performs better by classifying *DIE LINKE* while comparing with the *AfD* and the *UNION* and worse while comparing with *DIE GRÜNEN*. This result gives confidence and motivation to continue research in this direction, as well as to work on the improvement of the models.

In this thesis, we have looked at the problem of identification of the political leaning of news articles, often referred to as *media bias*. Most of the academic work here focuses on English-language media. We have developed a method to analyze German-language media. As shown in the Chapter 1 and Chapter 3, the shortcomings of existing methods are mainly due to two reasons:

- 1 The modeling of political media bias in computer science is very simplistic and does not make use of insights from the social sciences.
- 2 Creation of data sets to detect bias in the media are mainly based on three assumptions, which on the one hand are difficult to evaluate, and on the other hand often prove to be wrong, Section 3.4.

While the former reason is not the focus of this work, we were concerned with developing a method to overcome the latter problem. In our method, we make use of insights regarding phenomena such as *filter bubble* and *echo chamber*, manifested by the fact that users on social media tend to follow and share content that is consistent with their worldview. Recognizing a person's political orientation is not a trivial problem. We therefore restrict ourselves to politicians who belong to one of the parties that are represented in one of the state parliaments or in the Bundestag in Germany and whose political orientation can be derived from their party affiliation. We combine the theory regarding behavior on social media and the method for inferring the political orientation of a politician: a newspaper article is labeled based on the political orientation of the politician who also shared it. We use Twitter as the platform for collecting the data. To create the list of German MPs, we use AbgeordnetenWatch. As a platform for collecting user data, such as shared articles, we use Twitter. To extract text and meta information of the shared articles, we use the Python library *trafilatura* Barbaresi (2021). At the end of this process, we created a dataset consisting of 41,575 newspaper articles, containing raw article texts, titles, excerpts and meta information for the corresponding article, e.g., author and publication date. We assign a unique label to each article depending on which party a politician shared that article from. In the following, we summarize the designed research questions, describe and discuss the results of the selected approaches, and provide a brief overview of possible future work.

## 7.1 Discussion

We analyzed the created dataset with respect to three research questions:

- Q1.** How effective are standard classification approaches with frequency-based features and the standard neural approach in detecting bias from the right-wing political fringe, when using the articles shared by all other parties as negative samples?
- Q2.** How effective are the same standard approaches in detecting bias from right-wing political fringe, when training against the articles shared by other parties separately? As in the first research question, we label articles shared by *AfD* politicians as biased and the others as unbiased.
- Q3.** How do the same approaches perform when applying to the detection of the left-wing party *DIE LINKE*, i.e.: a) when using the articles shared by all other parties as negative samples, analogous to **Q1**; b) when training articles of *DIE LINKE* separately against articles of every other party, analogous to **Q2**.

For all three research questions we deployed standard feature based models, *Logistic Regression*, *Linear Support Vector* and *Multinomial Naive Bayes* and a state-of-the-art *BERT* model. We trained the feature based models based on TF-IDF vectors of excerpts *n-grams*(1–3) and fine-tuned the neural model on raw excerpts of news articles.

Across all selected models regarding **Q1** and **Q3a**, *BERT* achieved the best performance on over- and under-sampled dataset. From the results, we can see that the models definitely have difficulty distinguishing the biased class from the unbiased class according to our definition. This result can have many causes: (1) the class distribution. Since the models are trained to maximize a particular metric, it often leads to them simply choosing the majority class when predicting, and thus obtaining poor results on the minority class. Even the methods of under- and over-sampling lead to only marginal improvements in model performance, as both methods have their drawbacks. On the one hand, random under-sampling can lead to neglecting meaningful features of the majority class because they are not used in the training process. On the other hand, random over-sampling leads to duplication of randomly selected training examples of the minority class. Both methods lead to the manipulation of the original language usage to the detriment of the respective class, which is reflected in the resulting unsatisfactory model performance. (2) Mixing all articles shared by parties, except for *AfD* in **Q1** and *DIE LINKE* in **Q3a**, into one class, can potentially lead to the basic focus of each party, becoming mixed, and thus harder to distinguish from articles shared by *AfD* or *die DIE LINKE*, respectively.

The second hypothesis is in some way confirmed by the results of the experiments regarding **Q2** and **Q3b**. We see that the trained models perform significantly better here. In **Q2** and **Q3b** the models can distinguish the texts shared by the *AfD* and *DIE LINKE* significantly better from the texts shared by other parties when viewed in direct pairwise comparison with each other. Here again, the *BERT* model achieves the best performance in comparison to other models. Moreover, we observe a possible direct dependence between models' performance and actual proximity between political parties: in **Q2** we see, that the *BERT* model performs better in correctly classifying the articles shared by *AfD* when evaluating against *DIE LINKE* or *DIE GRÜNEN* and worse when evaluating against the more conservative party *UNION*. Similar to that, in **Q3b**

the *BERT* model performs better by classifying *DIE LINKE* while evaluating against the *AfD* and the *UNION* and worse while evaluating against *DIE GRÜNEN*.

Even though there is still much room for improvement, the results regarding **Q2** and **Q3b** show a positive tendency and confirm the assumptions we made at the beginning of the work to create the dataset: based on the theory of *homophily*, i.e., the axiom that similarity breeds connection Chun (2018), the already researched effects like *echo chamber* and *filter bubble*, which arise from the fact that people tend to spread content according to their understanding of the world, we label the shared content based on the party affiliation of the respective person who shared the content. Since we see in the results a fundamental confirmation of the basic assumptions made in this work to label the dataset, we discuss in the following possible further developments that can have a positive effect on the performance of the models.

## 7.2 Outlook

The amount of data plays a large role in interpreting the results of machine learning algorithms for classification. Since we have developed a methodology that is completely independent of any kind of manual annotation or interviews with experts, collecting new data is not a big hurdle. For this purpose, it would make sense to set up a system that continuously monitors the Twitter profiles of politicians and saves their tweets if they contain a linked newspaper article. Other data sources, such as Facebook, could also be used to monitor politicians' profiles.

Furthermore, the quality of the data could be improved. For example, with the help of stance detection, one could be sure that the tweets' content is congruent with the opinion of the article and does not criticize the article. This would prevent texts from being falsely assigned to a political direction and also contribute to improving model performance.

Another aspect that we think has potential is to select the one set of articles that depict exactly one topic on which all parties comment and share content. The topics that come into question are many and varied and are usually socially relevant over a longer period of time, such as immigration, the environment or foreign policy relations with certain countries with a tendency towards autocracy.



---

## Bibliography

---

- An, J., Cha, M., Gummadi, K., Crowcroft, J., and Quercia, D. (2012). Visualizing media bias through Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 6, pages 2–5.
- An, J., Kwak, H., Posegga, O., and Jungherr, A. (2019). Political Discussions in Homogeneous and Cross-Cutting Communication Spaces. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01):68–79.
- Baker, B. H., Graham, T., and Kaminsky, S. (1994). *How to identify, expose & correct liberal media bias*. Media Research Center Alexandria, VA.
- Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132.
- Barbarese, A. (2021). Trafilatura: A Web Scraping Library and Command-Line Tool for Text Discovery and Extraction. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 122–131. Association for Computational Linguistics.
- Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, 90(1-2):1–36.
- Baumer, E., Elovic, E., Qin, Y., Polletta, F., and Gay, G. (2015). Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics: human language technologies*, pages 1472–1482.
- Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.
- Bernhardt, D., Krasa, S., and Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5-6):1092–1104.
- Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Leveling, J., Monz, C., Pecina, P., Post, M., Saint-Amand, H., et al. (2014). Findings of the 2014 workshop on statistical machine translation. In *Proceedings of the ninth workshop on statistical machine translation*, pages 12–58.

- Bucher, H.-J. and Schumacher, P. (2006). The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *The European Journal of Communication Research*, 31:347–368.
- Chen, W.-F., Al-Khatib, K., Stein, B., and Wachsmuth, H. (2020a). Detecting Media Bias in News Articles using Gaussian Bias Distributions. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, page 4290–4300.
- Chen, W.-F., Al-Khatib, K., Wachsmuth, H., and Stein, B. (2020b). Analyzing Political Bias and Unfairness in News Articles at Different Levels of Granularity. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, page 149–154.
- Chen, W.-F., Wachsmuth, H., Al Khatib, K., and Stein, B. (2018). Learning to flip the bias of news headlines. In *Proceedings of the 11th International conference on natural language generation*, pages 79–88.
- Chun, W. (2018). Queerying Homophily Muster der Netzwerkanalyse. *Zeitschrift für Medienwissenschaften*, 10:131–148.
- Conover, M. D., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political Polarization on Twitter. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press.
- D’Alessio, D. and Allen, M. (2006). Media Bias in Presidential Elections: A Meta-Analysis. *Journal of Communication*, 50:133 – 156.
- D’Angelo, P. (2018). *Doing News Framing Analysis II: Empirical and Theoretical Perspectives*. Routledge.
- De Saussure, F. (2011). *Course in general linguistics*. Columbia University Press.
- De Vreese, C. (2005). News Framing: Theory and Typology. *Information Design Journal*, 13:51–62.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Druckman, J. N. and Parkin, M. (2005). The impact of media bias: How editorial slant affects voters. *The Journal of Politics*, 67(4):1030–1049.
- Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173.
- Fan, L., White, M., Sharma, E., Su, R., Choubey, P. K., Huang, R., and Wang, L. (2019). In Plain Sight: Media Bias Through the Lens of Factual Reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 6342–6348. Association for Computational Linguistics.



- Field, A., Kliger, D., Wintner, S., Pan, J., Jurafsky, D., and Tsvetkov, Y. (2018). Framing and Agenda-Setting in Russian News: a Computational Analysis of Intricate Political Strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3570–3580. Association for Computational Linguistics.
- Freitag, J., Kerkhof, A., and Münster, J. (2021). Selective sharing of news items and the political position of news outlets. *Information Economics and Policy*, 56:100926.
- Ganguly, S., Kulshrestha, J., An, J., and Kwak, H. (2020). Empirical Evaluation of Three Common Assumptions in Building Political Media Bias Datasets. In *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, ICWSM 2020, Held Virtually, Original Venue: Atlanta, Georgia, USA, June 8-11, 2020*, pages 939–943. AAAI Press.
- Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1):35–71.
- Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.
- Gentzkow, M., Shapiro, J. M., and Stone, D. F. (2015). Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1, pages 623–645. Elsevier.
- Gilens, M. and Hertzman, C. (2000). Corporate ownership and news bias: Newspaper coverage of the 1996 Telecommunications Act. *The Journal of Politics*, 62(2):369–386.
- Groeling, T. (2013). Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news. *Annual Review of Political Science*, 16:129–151.
- Groseclose, T. and Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.
- Hamborg, F., Donnay, K., and Gipp, B. (2018). Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, pages 1–25.
- Hamborg, F., Zhukova, A., and Gipp, B. (2019). Automated Identification of Media Bias by Word Choice and Labeling in News Articles. In *19th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*, pages 196–205. IEEE.
- Harcup, T. and O’neill, D. (2001). What is news? Galtung and Ruge revisited. *Journalism studies*, 2(2):261–280.
- Herman, E. S. (2000). The propaganda model: A retrospective. *Journalism Studies*, 1(1):101–112.
- Iyyer, M., Enns, P., Boyd-Graber, J., and Resnik, P. (2014). Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122.
- Kiesel, J., Mestre, M., Shukla, R., Vincent, E., Adineh, P., Corney, D., Stein, B., and Potthast, M. (2019). Semeval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839.

- Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., and Karahalios, K. (2017). Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2017, Portland, OR, USA, February 25 - March 1, 2017*, pages 417–432. ACM.
- Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., and Karahalios, K. (2018). Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal*, 22(1-2):188–227.
- Lim, S., Jatowt, A., Färber, M., and Yoshikawa, M. (2020). Annotating and Analyzing Biased Sentences in News Articles using Crowdsourcing. In *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 1478–1484. European Language Resources Association.
- Lim, S., Jatowt, A., and Yoshikawa, M. (2018). Understanding Characteristics of Biased Sentences in News Articles. In *Proceedings of the CIKM 2018 Workshops co-located with 27th ACM International Conference on Information and Knowledge Management (CIKM 2018), Torino, Italy, October 22, 2018*, volume 2482 of *CEUR Workshop Proceedings*.
- Lin, W.-H., Wilson, T., Wiebe, J., and Hauptmann, A. G. (2006). Which side are you on? Identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 109–116.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3111–3119.
- Milyo, J. and Groseclose, T. (2005). a measure of media bias. pages 1191–1237.
- Morgan, J. S., Lampe, C., and Shafiq, M. Z. (2013). Is news sharing on Twitter ideologically biased? In *Computer Supported Cooperative Work, CSCW 2013, San Antonio, TX, USA, February 23-27, 2013*, pages 887–896. ACM.
- Mullainathan, S. and Shleifer, A. (2002). Media bias. *SSRN Electronic Journal*.
- Mullainathan, S. and Shleifer, A. (2005). The Market for News. *American Economic Review*, 95(4):1031–1053.
- Oelke, D., Geißelmann, B., and Keim, D. A. (2012). Visual Analysis of Explicit Opinion and News Bias in German Soccer Articles. In *3rd International EuroVis Workshop on Visual Analytics, EuroVA@EuroVis 2012, Vienna, Austria, June 4-5, 2012*. Eurographics Association.
- Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543. ACL.

- Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., and Stein, B. (2018). A Stylometric Inquiry into Hyperpartisan and Fake News. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 231–240. Association for Computational Linguistics.
- Pryzant, R., Martinez, R. D., Dass, N., Kurohashi, S., Jurafsky, D., and Yang, D. (2020). Automatically neutralizing subjective bias in text. In *Proceedings of the aaai conference on artificial intelligence*, volume 34, pages 480–489.
- Puglisi, R. and Snyder Jr, J. M. (2015). Empirical studies of media bias. In *Handbook of media economics*, volume 1, pages 647–667. Elsevier.
- Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., and Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937.
- Recasens, M., Danescu-Niculescu-Mizil, C., and Jurafsky, D. (2013). Linguistic Models for Analyzing and Detecting Biased Language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pages 1650–1659. The Association for Computer Linguistics.
- Ribeiro, F. N., Lima, L. H. C., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., and Gummadi, K. P. (2018). Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. In *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018*, pages 290–299. AAAI Press.
- Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass communication & society*, 3(2-3):297–316.
- Spinde, T., Hamborg, F., and Gipp, B. (2020). Media Bias in German News Articles: A Combined Approach. volume 1323 of *Communications in Computer and Information Science*, pages 581–590. Springer.
- Spinde, T., Kreuter, C., Gaissmaier, W., Hamborg, F., Gipp, B., and Giese, H. (2021a). Do You Think It’s Biased? How To Ask For The Perception Of Media Bias. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2021, Champaign, IL, USA, September 27-30, 2021*, pages 61–69. IEEE.
- Spinde, T., Rudnitskaia, L., Mitrovic, J., Hamborg, F., Granitzer, M., Gipp, B., and Donnay, K. (2021b). Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing & Management*, 58(3):102505.
- Spinde, T., Rudnitskaia, L., Sinha, K., Hamborg, F., Gipp, B., and Donnay, K. (2021c). MBIC - A media bias annotation dataset including annotator characteristics. *CoRR*, abs/2105.11910.

- Stefanov, P., Darwish, K., Atanasov, A., and Nakov, P. (2020). Predicting the Topical Stance and Political Leaning of Media using Tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 527–537. Association for Computational Linguistics.
- Sunstein, C. R. (2001). Republic.Com. *Harvard Journal of Law & Technology*, Volume 14, Number 2.
- University of Michigan (2014). University of Michigan.: News bias explored—The art of reading the news (2014). <http://websites.umich.edu/~newsbias/>. Accessed: 2022-02-14.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.
- Wachsmuth, H., Kiesel, J., and Stein, B. (2015). Sentiment flow-a general model of web review argumentation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 601–611.
- White, D. M. (1950). The “gate keeper”: A case study in the selection of news. *Journalism quarterly*, 27(4):383–390.
- Williams, A. (1975). Unbiased study of television news bias. *Journal of Communication*, 25(4):190–199.
- Yardi, S. and Boyd, D. (2010). Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society*, 30:316–327.