Master Thesis

A News Recommendation Engine for a Multi-Perspective Understanding of Political Topics

Christina Luisa Kraus

Matriculation #: 376694

Supervisor: Professor Dr. Volker Markl
Advisor: Dr. Sebastian Schelter

31/07/2016

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Berlin, July 31, 2016

Christina Kraus

CONTENTS

List of Figures

List of Tables

1 English Abstract

People read news online rather than on paper more and more. In consequence, online news providers use recommender systems to improve the readers' experience. In reflection of today's fast-paced lifestyle, most of these systems focus on providing fast and easy access to articles, tailored on the user's interests. As a result, this may lead to a very superficial consideration of relevant political topics. Furthermore, news papers may be biased towards certain political directions.
We develop a model for a news recommendation system that yields insights to political topics from various perspectives. We propose a clustering approach on automatically extracted political features from news articles. We find that our approach outperforms TF-IDF feature extraction on political texts in consideration of silhouette coefficients. However, we face a challenge in the clear definition of political perspectives, which distorts external cluster validation. Nevertheless, the resulting recommendations are promising in regards to a multi-perspective understanding of political topics.

2 Deutscher Abstract

Der Anteil an Nachrichten, die online gelesen werden, ist im Vergleich zu Zeitungen in den letzten Jahren immens gestiegen. Nachrichtenanbieter reagieren auf diese Entwicklung mit Empfehlungssystemen, die sie auf ihren Webseiten integrieren, um das Leseerlebnis der Kunden zu verbessern. Diese Systeme legen besonderen Wert auf schnellen und einfachen Zugang zu Artikeln, die auf die Interessen der Leser zugeschnitten sind. Dies birgt jedoch das Risiko, dass insbesondere die Auseinandersetzung mit politischen Themen, sehr oberflächlich zu werden droht. Zudem besteht generell eine Gefahr, der politischen Befangenheit in der Berichterstattung.
Um dem entgegenzusteuern, entwickeln wir ein Modell für ein Empfehlungssystem, das aktuelle politische Themen aus mehreren Perspektiven beleuchtet. Automatisch extrahierte, politische Merkmalsvektoren, dienen als Grundlage für einen Clusteringansatz, der Artikel nach ihrer politischen Perspektive gruppiert. Mit diesem Ansatz erreichen wir, in Bezug auf Silhouettenkoeffizienten, bessere Ergebnisse als mit den populären TF-IDF-Merkmalsvektoren. Die Definition klarer politischer Perspektiven gestaltet sich jedoch als Herausforderung bei der externern Clustervalidierung. Dennoch, liefern die resultierenden Empfehlungen vielversprechenden Nutzen, hinsichtlich einer multiperspektivischen Betrachtung politischer Themen.

3 Introduction

Within only one year (2014-2015) the consumption of online news increased by ten percent. This is twice as much as the overall internet consumption [46]. As there are no printing costs for online news, publishers are able to provide constant updates on currently interesting topics. The latency between an event and its news coverage has shrunk to a minimum.

This development leads to a vast range of articles that one could read online as well as a very dynamic and fast-changing news environment. In order to guide their customers through this jungle of information, online news providers often offer additional functionality, which is known from E-Commerce stores and streaming websites, such as Amazon and Netflix: Recommender systems. Their goal is to improve customer experience and help customers find the most relevant items for them. For this reason, most recommendation approaches for news aim to tailor the news they display to the user's interest. However, this carries the risk of concealing information, as the customer might not see the full variety of the news offer. The term of the "filter bubble" describes this phenomenon, which is common in various domains of the internet age.

When it comes to political topics, news represent the communication instrument between politics and the people. This makes them a vulnerable information channel that should neither underlie concealment nor any other kind of bias.

3.1 Media Bias

Press is a very powerful instrument in a democratic state. It is not without reason that it is often referred to as the fourth power alongside with legislature, judiciary and executive [35]. It is *the* source of information for people living in a state and strongly determines the way they perceive current politics on a domestic as well as on an international level. Therefore, it is incredibly important to consider possible influences on this perception. Such perception distortions are called bias.

"It is well known that traditional media have a bias in selecting what to report and in choosing a perspective on a particular topic"[10]. McQuail [41] provides a more general definition of the media bias by describing it as "any tendency in a news report to deviate from an accurate, neutral, balanced and impartial representation of reality of events and social world." Entman [18] differentiates three types of biases. Distortion bias describes the phenomenon, when a journalist alters reality in his writing. Content bias is a lighter bias, in this case the journalist shows preference for one side or opinion in a conflict. At last, decision-making bias results from motivations and thoughts of the journalist.

Apart from differentiating different types of media biases, the reasons for manipulating press are manifold as well. While the previous differentiation refers to texts written by journalists and their conviction and mindset, Van Dalen [67] describe the phenomenon that some kinds of stories are just preferable over others, when it comes to publications in news. Such a bias is rather attributed to the publisher than to the writing of the journalist. They explain this phenomenon with the news value of power and conflict. The news value of power refers to the fact that powerful politicians are better represented in news. Powerful in this case means that their behavior has a strong impact on political outcomes. This also shows the vulnerability of news. They act as a political instrument, but at the same time have to fulfill a business goal for their publisher. The news value of conflict on the other hand refers to the fact that conflict between the executing part of the government and the parliament or weaker players in the opposition is also an attractive news story for publishers, especially when the opposition has the potential to influence political outcomes. In that case news are a very important communication channel for political stakeholders. Kaplan, Ethan and DellaVigna [26] show how strong the influence of biases may be within a real-life experiment in their research. They examine the effect of the media bias on voting in a real-life situation in the United States. A conservative TV channel raises the amount of votes for the Republican party in regions, where it broadcast. The effect could be measured as the channel was new and not yet available in all parts of the United States. This shows the danger in the underestimation of biased media, as political outcomes might be strongly influenced. The various definitions and explanations for the presence of media bias indicate that this is a human-made phenomenon. For this reason, Dallmann et al. [16] mention that it is important to automatically detect bias in news papers, as the other option of humans labeling texts carries the risk of not being objective either. Additionally, automatic detection enables applicability on large data sets. They develop methods for automated bias detection by taking into consideration four German news papers and examinining them in regards to biases. Therefore, they introduce coverage bias metrics and statement bias metrics. The former refers to if and how often German political parties are represented in the explored news papers by evaluating the frequency of the party name and popular representatives being mentioned in the article. Statement bias in contrast, is content-related and consists of a sentiment analysis that determines the mood that is conveyed in relation to a mentioned party. In a second step, it compares vocabulary similarity of the political manifestos of German parties and the vocabulary in the news. They reveal various ideologies in the vocabulary that appear in both, the manifestos and the article. Examples are freedom, solidarity and environment. This bears similarity to the Manifesto project and the Manifesto classifier, which we use in this work, as it classifies texts based on political manifestos that are labeled with ideologies.

3.2 Filter Bubble

In addition to the media bias, there is another source of influence on the way from a political event or action to the reader, the “filter bubble”. While media bias refers to the journalist or the news publisher, the filter bubble is a quite new phenomenon, born in the internet age. The term was coined by Eli Pariser [49]. It refers to the personalization of the internet. Nowadays, most websites, try to understand the needs and preferences of their readers and adjust their content accordingly. Even though this is helpful and time-saving, it leads to the

fact that everybody sees his own version of the internet, his filter bubble. By having preferences for certain topics, products or items in general, we enforce only these items to be displayed for higher user satisfaction. Facebook has become an important source for news, but it filters posts according to the user's behavior. This transformation of the timeline is not transparent [4]. Many users are not aware of the fact that Facebook holds content back. Facebook is a very famous and demonstrative example of the filter bubble.
Going further than social networks, the filter bubble is also a consequence of personalized recommender systems [10]. In order to make the content the user sees more meaningful to him and thereby increase the probability of shown items of any kind to be consumed, recommender systems narrow down the range of items that they display to the user and create a filter bubble for their users. Observing these developments, it seems to be only a matter of time until "the internet itself" becomes a restricted bubble for each person. Already, most people do not access websites directly, but search for whatever they are looking for in Google. Therefore, Google has an information monopole with a high risk for manipulation. The diversity of content that reaches a person online, is highly dependent on the algorithms that determine Google's search results.
The phenomenon of the filter bubble has been subject of research under various expressions. Others are tunnel vision or selective exposure. As the filter bubble is prevalent, its detection and analysis is not trivial. In the field of news, Tran and Herder [66] explore the filter bubble by comparing which events have been included in different timelines that show ongoing history of certain topics. They find that timelines differ drastically. A lot of them even neglect important dates. As a solution, they suggest to make readers aware of these missing points by linking to other timelines or aggregate a more complete one out of various timelines. Research on the effect of such a filter bubble shows that opinions become more extreme, when individuals holding these opinions are only exposed to them instead of a diversity of opinions. This is an undesirable trend as it may lead to political fragmentation. The trend increases with the possibility of news personalization, as people can customize news to their interest as well as opinion, while earlier, a broader audience had to be reached by the same media[20]. In their work they found that people differ in their desire to consume heterogeneous information. However, reasons for that have not yet been evaluated.
A way to overcome the filter bubble is to actively confront users with diverse information. Liao and Fu [33] do that in order to find out under which circumstances the user prefers being exposed to diverse information and when he prefers to be exposed to preselected information in the field of news. They consider a possibly related threat and personal topic involvement as factors that influence these preferences and find that low involvement in the topic leads to a preference in selective exposure, even if a threat is related, while high topic involvement increases the preference for diverse information. We assume that news recommendation systems that are trimmed on convenience and quick overviews, lead to lower personal topic involvement and reinforce selective exposure.

On the other hand, we even find approaches to overcome the filter bubble in the field of recommender systems. While users enjoy personalized recommendations according to their interests, as recommender systems achieve better targeting for their items it has become a new objective to bring new and surprising items to the user. This approach is very promising, as it might open up new high potential market areas. This surprising effect is known as "serendipity". However, recommending serendipitous items comes with the limitation of lower accuracy. It therefore opens up a big field of research that can also address the topics of media bias and filter bubble.

3.3 Approach and Thesis Outline

This work presents an approach for a recommendation system that has the aim to facilitate gaining a comprehensive picture on political events and decisions by recommending articles with differing and potentially controversial political perspectives on politics. We aim to explore ways to overcome the filter bubble and methods to unveil biased media.
The remainder of this work is structured as follows: In section 4, we concretise the problem to derive the hypotheses of this work. Then, in section 5, we present related work on recommender systems, the Manifesto corpus and previous approaches on identifying controversy in news. We also discuss the definition of political perspectives in this section. Section 6 describes our research approach, including the algorithms used and section 7 deals with its technical details. We present our data in section 8. Sections 9 shows the experiment setup, the results and their evaluation, before we conclude our work in section 10.

4 Problem Definition

Let us assume that lot of people read news articles to be kept up to date. They want to know what is going on in the world, improve their political understanding and be able to discuss political topics in everyday life. Thereby, they might or might not be aware of the possible effects of the filter bubble or biased media. There is a risk of being kept in one's bubble of ideologies. Twitter, for example, experiences clear network segregation regarding retweets. Political information of a certain perspective is mostly retweeted within a sub-network of that perspective rather than in an oppositional one [15]. Facebook profiles polarize in a similar way according to the users' political affiliation [4]. There are people that might not bring the desire to step deep into a topic or confront themselves with opposing opinions on their own initiative[33], but being confronted with homogeneous news can distort facts over time [29] and lead to more extreme opinions [20].
In order to overcome the two manipulating phenomena of media bias and filter bubble, it is helpful to read multiple articles on the same topic [50]. Some people might already try to diversify the sources from which they collect their information to cover their information need from various political angles. Online

initiatives, such as "Peace it together" promote such behavior through short movies that illustrate the benefits of diverse information sources [1]. Being aware of different views on political topics is a good knowledge base for building one's own opinion. Furthermore, reading different articles on the same topic may help to get a better feeling for subliminally carried opinions in seemingly objective news articles. However, finding relevant and diverse sources to gain an unbiased picture of political situations is time consuming, regarding the great choice of online news. Therefore, it is necessary to find a way to facilitate the access to diverse information.
In general, decreasing the customer's search costs bears great potentials in electronic markets[3]. Furthermore, customers are willing to spend more time if search costs are low [64]. Therefore, a recommender system that provides the readers with a collection of multi-perspective articles on political topics may increase the customers' willingness to invest more time in the actual reading and thereby help to deepen their political understanding. While recommender systems are already very common support instruments on news websites, they mostly focus on personalization techniques and performance issues. They aim to help the readers find articles that they might be interested in, in order to increase user satisfaction and thereby increase the conversion rate of read articles on the website [52]. For this reason, former research in news recommender systems also puts emphasis on these questions. As a result, there are well working approaches for this task, as it is possible to build user profiles, by deriving interests from previously read articles or to suggest articles based on general news trends [34]. Furthermore, there are approaches that aim to improve knowledge. They provide additional background information on the texts, for example through semantic annotations of text [25, 48, 12, 13]. This background information puts the article into a wider context and can increase the information transfer through reading. However, this might still not specify the political dimension of an article. Our goal is to point out different political perspectives on certain events by juxtaposing a small number of articles on the same topic. While this still increases reading time, it saves a lot of time as it would be hard for the readers to manually filter out articles that contain controversial political views from a big pool of news.
A first step towards such a recommender system is to find a way to politically characterize news texts. News articles with political profiles represent a valuable data representation for its development.
The research field of information retrieval deals with the analysis of texts and automated information extraction. Common techniques of text characterization are bag-of-words vectors and TF-IDF vectors [2]. The latter represents a popular method for topic extraction from unstructured texts. Furthermore, sentiment analyses are a way to analyze texts beyond their literal content and find out whether the text conveys positive or negative sentiments. Still, these methods do not seem sufficient to extract a political viewpoint from an article

[1] http://peaceittogether.com/learn-more/blog/diversify-your-media-sources-#.V5xU1riLQ1I

text. To work on such problems, there is a domain-specific classification approach that aims to politically classify any given text onto a political scale of left and right. The classifier has been trained on political manifestos that have been labeled with political ideologies. Apart from a clear class assignment, it provides predictions for these ideologies upon which we build our model. Within this work, we aim to answer the following questions:

1. **Is it possible to automatically derive political perspectives from news texts using machine-learning algorithms?**
   (a) Is the Manifesto classifier an appropriate instrument to politically characterize texts as a basis for a recommendation engine?
   (b) Does text characterization with the Manifesto classifier outperform traditional information retrieval methods, such as bag-of-words or TF-IDF?

While clustering on TF-IDF vectors is quite common and has been examined before, the Manifesto classifier quantifies the text into ideological dimensions. This opens up a new space concept of political ideologies. We aim to explore this space and as we intend to group news articles based on their political perspectives, we apply a clustering approach on this new space. Therefore, we address the following question:

2. **What clustering approach is suitable to group news articles based on their political dimensions?**

Answering these questions lays a foundation for a news recommendation approach that motivates people to spend more time reading and examining political topics in greater depth and from different political perspectives. It therefore has the potential to deepen political knowledge and enforce political education.

5 Related Work

In this section, we provide the theoretical background for this thesis. We first give an overview of recommender systems in general, by going into different functions and categories of current systems. Then, we differentiate the two most common recommendation approaches, namely: Collaborative filters and content-based recommendations. As the topic of news recommendation has been treated in former research, we discuss existing goals and approaches in order to set them in relation to our own problem definition and the resulting approach. In the implementation of our approach we use a text classifier that is based on a document corpus containing political manifestos. Therefore, we introduce the classifier and the corresponding document corpus. We also discuss other approaches that aim to identify controversies in news. At last, we present ways to differentiate political perspectives.

5.1 Recommender Systems in General

"Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user" [52]. This is a rather general definition of a recommender system. Based on this definition, the main goal of recommender systems is to reduce choice in an environment of information overload and thereby increase the conversion rate. In an online store this can mean to increase sales, on a streaming platform such as Netflix, it is equivalent to a user choosing and watching a recommended movie. In our case of news recommendation, items refer to news articles and users to news consumers. Consequently, the conversion rate would relate to a news consumer choosing a suggested news article to read.

In regard to this definition and the corresponding goal of recommender systems, multiple recommendation approaches have been implemented that differ in many aspects depending on the domain in which they act, the nature of the items and users as well as the strategy used to find the most suitable item for a user. A first and very general differentiation criterion for recommender systems is personalization. A common representative of an unpersonalized recommendation approach is a top seller list [21]. Top seller lists rely on the fact that popular items are more likely to be purchased by customers than a randomly chosen item [57]. In contrast to top seller lists, personalized recommendations take into account some kind of user profile and adapt their recommendation to it. While top seller lists are easy to implement and often successful they do not always satisfy the providers' needs. Providers may have greater benefits in recommending items that are less popular as they might be cheaper to acquire and provide a greater profit margin [23]. These less popular items are called niche products. An example for this phenomenon are movie streaming platforms, such as Netflix. They face higher costs when playing popular movies. Therefore, it is in their interest to find niche products to recommend to their users. On the other hand, the interest for niche products can also be on the side of the user [11], as he might either already know about popular items or can easily find this information online. Niche products usually relate to the diversification of the product offer. Adapting this concept to news, we need to diversify news sources and underlying political perspectives. Finding niche items is often time consuming and therefore related to high search costs, which many people are not willing to pay. This is a starting point for other and particularly personalized recommendation approaches. While the logic for top seller lists is rather trivial, matching niche products to user profiles is a difficult task and therefore a way for providers to stand out from the competition. This explains why research activities mostly focus on the manifold implementations of personalized recommendations [52]. Apart from recommending individual items, recommender systems may also have the function to assist in suggesting item bundles or sequences. Showing complementary news articles to an article a user has already chosen is a type of item bundle - our recommender has the function to bundle politically controversial articles in order to widen the readers perspective.

Looking at personalized recommender systems, there are two main recommendation techniques: Collaborative filters and content-based recommender systems.

5.1.1 Collaborative Filters

In commercial use, collaborative filters are the most popular approach for recommendation generation [52]. The idea of collaborative filtering is to compare user preferences for items. Examples for preference representation are purchase histories, click streams as implicit or ratings as explicit preference expression. Users with similar preferences are considered to be similar and therefore to like the same items. For a new user, the system looks for similar users and recommends an item that has not yet been purchased by this user but by similar users. The most common approach to solve this problem is the nearest neighbour approach. Besides neighbourhood approaches, latent factor models build a group of collaborative filters. Through matrix factorization they aim to identify latent features to explain user preferences. To achieve this they project both, users and items to a joint latent factor space [28]. Furthermore, Bayesian networks, clustering and horting are other instruments for the implementation of collaborative filters. Collaborative filters have the great advantage that no additional information on neither users nor items is necessary to express a recommendation. However, they come with some limitations that should not stay unmentioned: Scalability is an important challenge in the implementation of collaborative filters as they need to search tens of millions of neighbors and this number will be increasing with every new user as well as every new item.
The second big challenge is sparsity. In usual settings and as the amount of products increase we have matrices with thousands or millions of possible items of which most users have only rated a fraction. This not only leads to computationally costly calculations, but also to poor recommendation quality [56]. To avoid these limitations, Sarwar et al. [56] introduce item-based collaborative filtering. Instead of comparing users and their preferences it puts the focus on the items and the relationship between them. Another shortcoming of collaborative filters is the cold-start or first-rater problem [58]. A user that has not rated any item yet cannot be compared to the rest of the users in order to find similar users. Correspondingly, an item that has not been rated by any user cannot be recommended to any user. This is particularly important regarding news recommendation as new news articles constantly appear and the recommendation of an article even in a highly frequented environment such as Google News would take hours [34].

5.1.2 Content-Based Recommender Systems

Content-based systems match items to users based on item features and user interests. A content-based recommender system takes into account documents or item descriptions that a user has previously rated. The system uses this information to extract item features and build a user profile [44]. Lops et al. [36] introduce a high-level architecture for content-based recommender systems: To

express a recommendation, a content-based recommender system follows three steps. At first there is the content analyzer. This part of the system extracts features from items. In a second step there is the profile learner. Based on all items that a user has consumed/purchased in any way, it generalizes and builds a user profile. At last, the system matches the determined profile to the available items using similarity measures, such as cosine similarity. Depending on the availability of information, recommender systems can involve a feedback loop that enable them to learn over time depending on how previous recommendations have been handled by the user [1]. The systems differ in the way they build their user profiles. Many systems use information retrieval methods, such as vector space models to represent items [32]. Another option is to take into account external data, for example Linked Open Data, in order to include some semantic analysis of the content [17]. Content meta data is another possibility for additional information. An important advantage of content-based recommender systems is that it reduces the cold-start problem because recommendations only rely on the active user and his user profile. Similarly, items that have not yet been rated can also be recommended as soon as their features are defined [9, 36]. Another advantage that may raise acceptance among users is the fact that content-based recommender systems are transparent and the user is able, if the information is provided, to understand how the recommendation appeared. This is not the case for collaborative filters where a user is compared to all other users in the system and totally unrelated items may be recommended [36].
A common limitation of content-based systems is a lack in domain knowledge. Domain knowledge is necessary to define and extract meaningful item features. Therefore, domain experts often have to manually label the data. A further, often mentioned limitation, is the so-called "serendipity problem". It refers to the fact that content-based recommender systems tend to recommend very similar items over time as they provide similar item features. While the cold-start problem or first-rater problem is reduced for new and unrated products, new users will still have to rate some items before they will receive a good recommendation[36].

5.2 News Recommender Systems

As mentioned previously, collaborative filters have proved very successful for recommendations in online stores like Amazon or streaming platforms for videos and music, like Netflix or Spotify. However, the general "rules" do not directly apply in the field of news. This is due to different reasons: First of all, there is a difference in the users' aim, when coming to a news page. Users in this field are generally not looking for something specific. They rather want to be updated on their fields of interest [34]. This means that in news the user does not have a specific information need, it is novelty that makes a story interesting. Also, the choice of an individual article cannot be considered independently. It often depends on what the user has read before, he might want to follow up on certain topics. Then, news are very dynamic. The popularity of certain topics changes frequently. Recency therefore is a major aspect. Whenever something new and meaningful happens, old topics might suddenly experience a huge loss of popularity [31].

5.2.1 Goals of Former Research On News Recommender Systems

Specific news recommendation approaches have been in research focus under various aspects. Apart from using different recommendation approaches, like the ones described in the previous section: collaborative filtering, various content-based approaches or hybrid ones, their actual aim varies as well. There are systems specialized on specific topics, systems to speed up news consumption, other systems to deepen knowledge through news consumption and so on. They refer to make reading news more convenient. They tackle questions on how to find the most important articles without scrolling through pages of irrelevant news, how to perform this in reasonable time and how to take user interests into account, rather than how to deepen knowledge in a specific domain. To give an overview on former research in this field, we divide the approaches by their key aspect. In total, we identify three groups of recommendation approaches that focus on the following aspects: Personalization, content and recency.

1. **Personalization:** They try to best fit the reader's interest.

2. **Content:** They try to semantically analyze the content in order to derive additional information.

3. **Recency:** They tackle the news specific issue of a very dynamic environment, with continuously new topics/articles popping up.

Besides these three groups, there is another important aspect that is present in all recommender types: performance in terms of speed and scalability. This is a cross-sectional topic that gains in importance with the increasing amount of available information. Unlike online shops with rather small product ranges, for news the performance aspect should be considered quite early due to the dynamic nature of the field. Even though some approaches put their main effort in improving the scalability of a recommender system, it can actually not be considered as an individual goal as it is always connected to a certain recommendation approach and all recommender systems used in the online environment should be scalable to be able to be used in real-life scenarios. Similar to the Netflix-competition [5], which aims to improve recommendation quality by providing a realistic data set and high incentives, there is the yearly held "CLEF-NewsREEL" competition particularly for news [27]. It deals with the specific character of the news environment with quickly and frequently changing sets of users and items. In form of living labs, it provides a platform to compare news recommendation approaches in realistic scenarios.
In the field of personalized news recommendations, we find the following approaches: One of the first systems that can be considered a news recommender system was called "NewsDude" and published in 1999 [7]. At that point of time reading news online was not common. Its area of application is the car radio.

It enables people to only listen to what they are interested in. Therefore, it requires explicit feedback from the user. The user has to specify whether he is interested in the article, whether he already knew it, whether he wants more information about that topic and whether he needs some deeper explanations. Furthermore, the system takes into account to which proportion of a story a user listens, as an implicit feedback. In a continuation of this work, Billsus et al. [8] develop a personalized recommender system for wireless news access. Apart from facilitated access to news, this approach tries to reduce costs, as wireless web access at that time was costly. Its recommendation technique is rather basic. An article receives a positive rating, whenever a user selects it. However, similar articles to previously read articles are not recommended as the authors assume that once the information has been consumed, the information need is fulfilled. The system differentiates short-term and long-term user interests, where short-term represents current news topics and their follow-up stories and long-term interests are rather general fields of interests, such as technology or science. The short-term interest model is content-based and uses a nearest neighbor text classification approach with a vector space model and TF-IDF weights. The long-term model, models predictions for new and unrated topics with a naive Bayes classifier based on previously identified user preferences. These preferences are represented as informative words within news articles that are related to each user.
Liu et al. [34] analyze the users click behavior in order to make personal recommendations of news articles to the user. They want to serve the user with the most interesting news for him. Therefore, they use a Bayesian model that takes into account user interests as well as general news trends. The recommendation technique is hybrid. The system automatically creates user profiles based on the users interaction with the system and additionally derives general trends from aggregated user actions.
The "SCENE" approach tackles the challenge of personalization as well as scalability. It is a hybrid approach considering the content of articles in a first step as well as user profiles based on accessed content, access pattern and preferred name entities in a second step. Based on their content, articles are clustered into news groups that are then matched with the topic distribution in a user profile. Their model is a budgeted maximum coverage problem, which implies that the selection of one news item has an impact on the selection of further items in the future [31].
Approaches that focus on the content aim to semantically analyze an article for various reasons. For example, Lavrenko et al. [30] suggest a specialized approach that targets to identify articles that influence the financial market. Their analysis solely focuses on the content of the article and not on the reader. To achieve their goal they use Bayesian language models to classify texts into trend groups. These Bayesian models estimate the likelihood of a trend based on the word distribution in the document. Ingvaldsen et al. [25] follow a general approach. Their content analysis helps the users to better understand the context of an article and deepen their knowledge beyond the actual text. To achieve this, they semantically enrich news texts by providing information on

certain included entities, taken from WikiData. Recommendations for articles take into account the geographic location of the users, personal interests and time. However, these are optional. The user has the possibility to activate or deactivate the information used for his recommendations. User interests also arise from WikiData entities in articles. By using the Apache Spark engine to stream RSS feeds, the approach also considers the scalability challenge of news recommendation. A similar approach has been introduced by [48].
In two consecutive works Capelle et al. [12] develop semantic-based news recommendations. In a first approach they analyse news content with a SF-IDF weighting scheme. In addition to traditional TF-IDF measures, SF-IDF also considers the actual meaning of the terms. A follow-up work [13] goes further than this. The Bing-SD-IDF+ news recommender follows a hybrid approach. A user profile that contains information about previously read articles is matched to new articles with the help of two similarity measures. These calculate similarities between named entities and between synsets. A synset represents a term with semantically equal terms, taken from a semantic lexicon.
The following approaches aim to handle the dynamics in the news environment. In contrast to online shops, news are constantly updated. Owen Phelan et al. [47] use Twitter to figure out current topics of interest in real-time and look for them in RSS feeds. They match the topics from Twitter with the RS-Feeds by evaluating the co-occurrence of popular terms.

5.2.2 Existing approaches on the market

Apart from research approaches there are a couple of publicly available approaches on the market that pursue similar objectives as we do in this work. An example is the German project "Pressekompass"[2] that has been established by the company Opinary in cooperation with two very popular news providers, the newspaper "Die Welt" and the online news platform "Spiegel Online". The page displays a question on a controversially discussed news topic, for example whether the commotion regarding the panama papers is justified and whether offshore companies should be forbidden. By means of a compass the page visualizes different answers to that topic and how certain news papers positioned themselves in regard to these questions. In the center of the compass there is an avatar, which allows the user to take a position on this question as well. In addition, it is possible to read facts about the questions that explain how it evolved. Figure 1 shows the "Pressekompass" for the exemplary question on the panama papers.

[2] http://pressekompass.net/

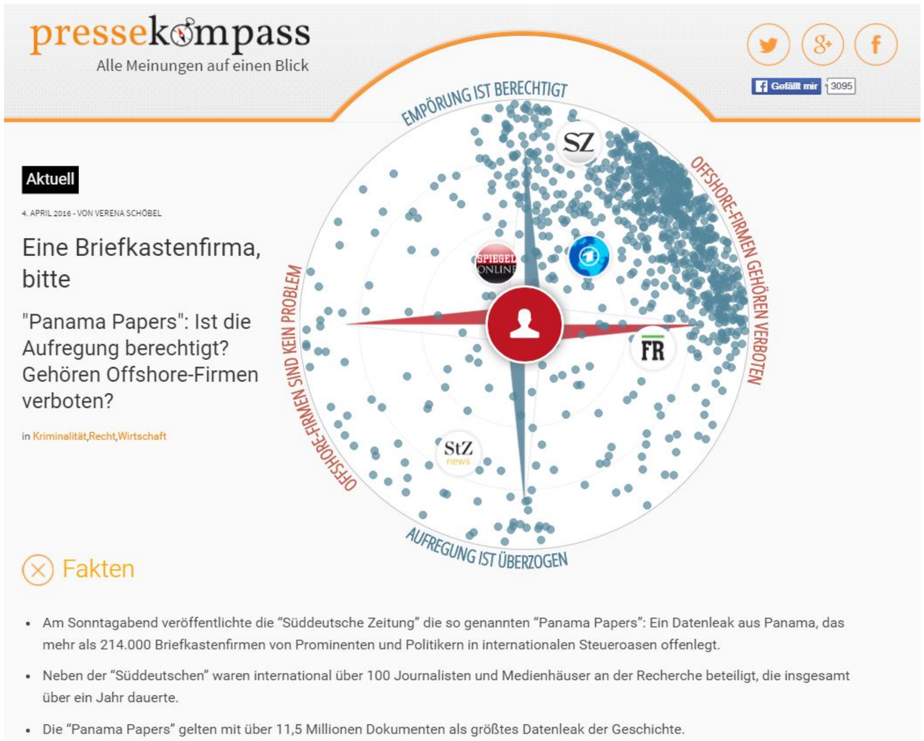


Figure 1: Pressekompass on Panama Papers[3]

The same company that developed the "Pressekompass" has established another instrument for online news. It is a two-dimensional poll, which is displayed underneath some articles on online news websites, showing two opposing opinions on the corresponding topic. Apart from the actual article, the poll positions some more articles or opinions of politicians on the field and as before, offers the reader the possibility to position his own avatar to demonstrate his own perspective. Figure 2 shows such a poll on the discussion of safe countries of origin.

[3]Source: http://pressekompass.net/panamapaperskompass/

Figure 2: Opinary Poll on Safe Countries of Origin[4]

Another initiative from the German "Federal Agency for Civic Education" maps political opinions of users to German political parties in order to show, which political party best reflects one's opinion in an upcoming election. Different to what has been discussed earlier, this initiative requires active involvement of the user. He has to express his opinion on a number of statements that are related to currently relevant political questions. In a next step he decides which of the statements are particularly important to him. Dependent on his choices and the political parties he is interested in, he receives values of accordance to each of these parties[19]. Figure 3 displays the user interface of the "Wahl-O-Mat" and a result view. However, this application is not directly connected to news events, but rather to political positions regarding the upcoming electoral period.

[4]Source: www.welt.de/debatte/kommentare/article156184843/Nein-liebe-Gruene-Migration-ist-kein-Menschenrecht.html

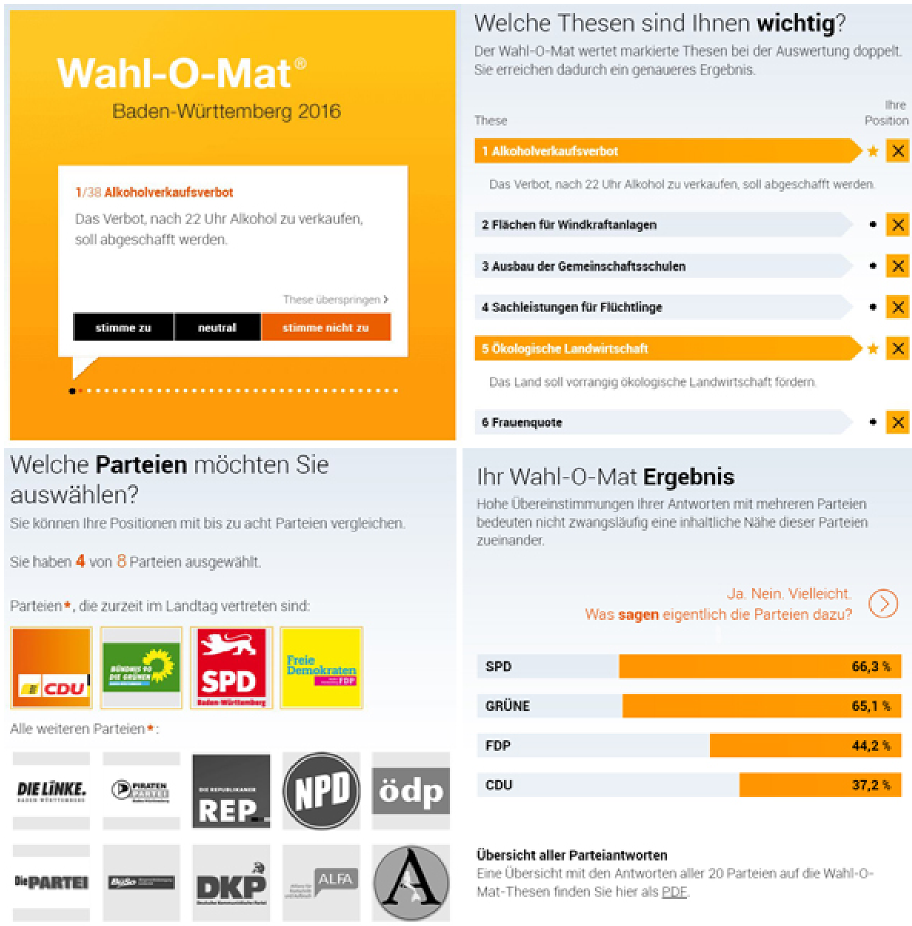


Figure 3: Exemplary four steps of the "Wahlomat"[5]

5.3 The Manifesto Corpus and the Manifesto Classifier

Content-based recommender systems frequently use classification approaches to build user profiles. Classification is defined as "the construction of a procedure that will be applied to a continuing sequence of cases, in which each new case must be assigned to one of a set of pre-defined classes on the basis of observed attributes or features" [43]. To make user profiles comparable it is helpful to categorize items into predefined categories and then derive user interests in these categories rather than single items. There are various options for the categorization of news. High-level topic representations, such as technology, politics, economy, culture, sport and entertainment like in "Google News" or "Spiegel Online" are very common and frequently used. Regarding the subtopic

[5]Source: www.wahl-o-mat.de/bw2016/main_app.php

of politics, political perspectives are another way to categorize articles. The task of classification needs a training set, on which the classifier is trained to make its choices. For text classification this is usually some document corpus. The “Manifesto project” provides such a corpus called the “Manifesto Corpus”. The corpus contains more than 1800 documents from 40 countries. These documents are electoral programs of political parties starting from 1945 to date. The documents are hand-annotated and assigned to 57 categories that represent different policy goals [42]. Examples for these goals are “Centralisation”, “Controlled Economy”, “Democracy”, “Military” and “Peace”, often combined with positive or negative sentiment as this may imply different meanings. The corpus therefore is a valuable resource for all kinds of policy-related research. One example is a multinomial logistic regression classifier that aims to predict political affiliation [6]. We use this classifier for our approach and refer to it as the Manifesto classifier from now on. For now, the classifier only works on German texts, but it might be extended to a multi-language version. Its goal is to classify a given text on a political left-right scale This is done by transforming the given text into a TF-IDF normalized bag-of-words vector. The classifier then classifies the text with a multi-class logistic regression, by comparing the vector to the vectors of the document corpus. In the original version the classification categories were the factions of the German “Bundestag”. The work has been extended to the more comprehensive Manifesto corpus.

5.4 Identifying Controversial Points in News

The aim of our work is to find a way to deepen political knowledge and understanding when reading news. Therefore, it is necessary to recommend the appropriate items to the users based on what they are already reading. We discuss recommendation techniques and applications of news recommender systems in the previous sections. We also introduce the classifier, which we aim to use to find controversial texts. However, in this last step, we present some alternative ways that have been used to identify controversial points in news in the past that are not necessarily related to a recommender system.
One way to find out whether a news article is controversial or not is to analyze the corresponding comments that were made by its readers. Lourentzou [37] conduct this analysis with a sentiment analysis tool that is specialized on social media texts. They expect controversial topics, to have particularly opposing sentiments in their comments, very positive and very negative ones. In contrast to our aim, that is to find multiple articles that represent the controversy of a topic, this application highlights the most controversial sentences for every read article and shows additional relevant tweets. Likewise, Choi et al. [14] assume that controversial topics raise contradicting reactions. They further state that most controversial topics, on a higher level, have some related subtopics. Each of them may correspond to a different sentiment, which explains the controversy of the main topic. Based on a search query containing a topic, they retrieve multiple result documents and analyze whether and how much they contradict each other. In a next step they identify corresponding subtopics, that actually

raise the controversy. In an application example, the authors show a high level topic, with multiple subtopics that carry different sentiments.

5.5 Political Perspectives

In order to develop a system that enhances diverse views on political topics, it is necessary to define possible representations of these views. Therefore, it is important to understand that political journalism is linked to democracy. "Journalists provide the information on which citizens will be able to judge between competing candidates and parties[40]. This implies that there is a link between political perspectives of news and the political parties of a country. However, not every party represents their own perspective.
Among democratic states there are different political systems: one-party systems, two-party systems and multi-party systems [68]. Within these political systems, we find political parties that incorporate different ideologies. While there are no strict typologies of political parties [38], we find three main families of parties [68]: liberal parties, socialist parties and conservative parties. Besides, there exist subfamilies, such as anarchists, communists or green/ecological parties. The Manifesto classifier abstracts this to only two political perspectives, left and right. All these divisions display possible groupings for the news articles in our work.
As we develop our model on German news texts, we consider the German party landscape in more detail. Figure 4 shows the current distribution of seats in the German Bundestag. However, there are more parties that are part of state parliaments like the liberal party FDP or the right-wing populist party AFD or the extreme right-wing party NPD.

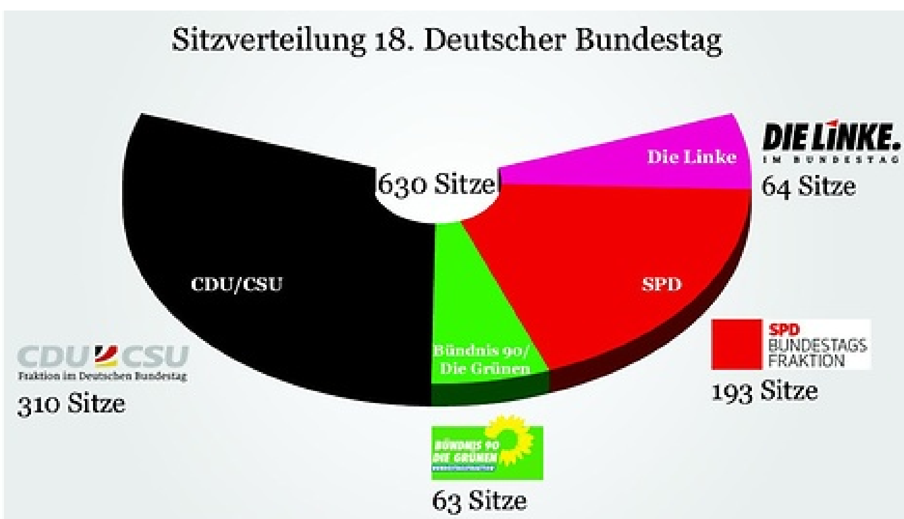


Figure 4: Distribution of Seats in the German Bundestag[6]

6 Our Approach

This section presents the methodology we apply to answer the research questions raised earlier in this work. We evaluate and transfer the methods we described in the *Related Work* (5) section to our problem and develop a comprehensive approach for a news recommendation engine of controversial political news articles. On a general level the aim of our recommendation approach is the following:
Group a set of news articles on the same topic into groups of political perspectives and choose a representative article for each group to show to the user.
Previous research on recommender systems has shown that different domains have different requirements and possibilities in terms of recommendation approaches. For example, stores that have a huge range of products and product categories may experience immense benefits from comparing customers' buying histories across different product categories, without taking into account the actual content of the products. In the context of books or movies this may not be the case as these items have to satisfy very individual tastes and interests. Their acceptance by customers almost solely depends on their content. News articles not only base on their content, but pose a further challenge as they are mostly available online and without cost. Therefore, customers usually do not have user profiles that offer historical information on their behavior. These conditions clearly favour a content-based recommendation approach. And while traditional content-based recommender systems have the limitation that they require domain specific expert knowledge, news articles consist of text, which offers manifold possibilities of automated analysis and feature extraction. New, incoming articles can run through an automated text categorization process, so that the articles reach the application with minimal latency. Thus, a content-based approach with automated feature extraction is also not subject to the cold-start problem, as new items are automatically characterized, when they come in. Another, often mentioned limitation to content-based recommender systems, is the so-called "serendipity problem". It refers to the fact that content-based recommender systems tend to recommend very similar items over time as they provide similar item features. However, Matt et al. [39] suggest that this limitation depends on the market type and the recommendation technique. They find that content-based systems increase novelty and serendipity in the case of blockbuster markets, but not in niche markets. In the news environment, news headlines could be considered as blockbusters. Headlines are usually connected to short, summarizing articles on the main facts of a topic. A content-based recommendation approach is therefore appropriate to find articles that treat similar topics, but highlight them from a different perspective. To support novelty and serendipity of recommended items, we characterize an article beyond its literal content. Looking at a news article, the content could be some facts on the "European refugee crisis", such as the closure of the Balkan

[6] Source: www.bundestag.de/bundestag/plenum/sitzverteilung_18wp

route. Figuring out the political orientation goes further, by asking how the article reports on this event. Does it utter appreciation or criticism? Thereby, it is important to ensure, that a potential user would not receive duplicate information but recognizes the additional insight on a topic.
A high level architecture for a content-based recommender system consists of a *content analyzer*, a *profile learner* and a *filtering component* [36]. In the following, we elaborate these components to our problem definition.

6.1 Content Analyzer

In the content analyzer "Data items are analyzed by feature extraction techniques in order to shift item representation from the original information space to the target one"[36]. Referring to our approach a data item corresponds to the text of a news article. A simple representation of this information is a space with all possible words, in which an item is a vector consisting of the counts of the words in an article text - a bag-of-words representation. While this is a quite accurate representation of a text, it does not cover further concepts within the text. The same text described with synonyms would lead to a very different bag-of-words vector. Furthermore, such a vector can be very high-dimensional and sparse, which causes challenges in the later evaluation. We aim to customize the analysis to political texts and therefore use the Manifesto classifier to project this information into the target space. The original target space is a 57-dimensional space, with each dimension representing a political idea or ideology. The classifier provides functionality to reduce this space to only two dimensions: left and right. The Manifesto classifier is trained on labeled political manifestos and primarily provides us with a 57-dimensional normalized weight vector that contains one value for each underlying label. In a next step, it maps the labels to a political "left-right" scale. As a reference corpus the Manifesto classifier uses 14 political manifestos from 2005 to 2013, published by the following German political parties:

- Alliance'90/Greens (2005, 2009, 2013) → Grüne
- The Left. Party of Democratic Socialism (2005) → PDS
- The Left(2009, 2013) → DIE LINKE
- Social Democratic Party of Germany (2009, 2013) → SPD
- Free Democratic Party (2013) → FDP
- Christian Democratic Union/Christian Social Union (2005, 2009, 2013) → CDU/CSU
- Pirates (2013) → Piratenpartei
- Alternative for Germany (2013) → AFD

These manifestos have been split into sentences that are labeled with the 57 possible labels. Table 1 shows all these labels.
The classification process works as follows: In a first step, it splits sentences into their single word components and corresponding counts of words. Then, in a second step the classifier transforms this word-count matrix into a normalized TF-IDF vector. TF-IDF has the advantage that it assigns weights to the terms depending on their relevance for the text. The TF-IDF vector is constructed as follows [55]:

$$tf(t,d) = \frac{f(t,d)}{max\{f(w,d) : w \in d\}} \tag{1}$$

Equation 1 measures the term frequency of a term t in a document d, normalized by the maximum term frequency of any term in that document.

$$idf_i = log\frac{N}{n_i} \tag{2}$$

In equation 2, N is the total number of documents and n is the number of documents in which the term occurs.
In the Python implementation the final TF-IDF formula has the following form [60]:

$$tf * (idf + 1) \tag{3}$$

The third step is the logistic regression classification. It uses the following sigmoid function (Equation 4) to estimate probabilities for each label [70]:

$$Pw\,(y = \pm 1|x) \equiv \frac{1}{1 + e^{-yw^T x}} \tag{4}$$

with: $x \rightarrow$ data, $y \rightarrow$ class label, $w \rightarrow$weight vector.
As we have multiple classes, the classifier returns probability estimates for each class. After its training, the classifier runs through these steps for each new incoming document and predicts its labels and the political direction based on the reference corpus. We use the two-dimensional representation of the political direction in a first step and later work with the original Manifesto label vectors. In this context, we intend to find out, whether the Manifesto labels are a suitable way to classify texts based on their political perspectives. For validation reasons, we compare our method to two simpler, but very popular approaches that we use as a baseline. We compare the Manifesto labels as a set of clustering features to a simple bag-of-words vector (referred to as BOW from now on) and a TF-IDF-vector. BOW is one of the simplest ways to quantify a text. It splits the document into its words and the corresponding frequencies. A vector is as long as the whole vocabulary of the document corpus and contains frequencies for every word. It represents our original information space as a whole and does not extract any additional information from the data. The TF-IDF-approach, as explained earlier in this section, is more elaborate. It is frequently used for topic extraction from document corpora [62]. We already input documents of the same topic. Therefore, we assume the comparison of the vectors to express more

subtle differences within the texts. Also, political perspectives are probable to focus on different aspects of a topic. Nevertheless, similar to the BOW, TF-IDF vectors do not capture deeper concepts than words, so we expect to reach the best characterization with the Manifesto classifier.

6.2 Building Political Profiles through Clustering

The next step in a "traditional" content-based recommender system is the profile learner. While this usually refers to user profiles, based on user interactions with the system, we build political profiles for articles. To construct such a profile, the traditional profile learner uses machine-learning methods to generalize from specific user actions to a profile of interests [36]. Similarly, we generalize from individual label vectors for each article text to groups of vectors and texts, respectively, that represent political profiles. We do not aim to adapt the content to a specific user as political perspectives should be universally valid. The recommendation is solely content-dependent, not personalized.

Again we follow a two-step approach. In a first step, there are only two options for political profiles: politically left and politically right. The Manifesto classifier can classify each article into one of those two categories. In the second step, we consider the more complex article representations of the Manifesto label vectors, the TF-IDF vectors and the BOW vectors. As all these representations are rich in features, a generalization requires further processing. Regarding machine-learning methods, there are different ways to decide on the profile assignment: Supervised and unsupervised learning methods. While supervised methods, such as classification require a prior definition of the different profiles, unsupervised methods like clustering are more exploratory and therefore allow for more flexible models [51]. It may be difficult to generalize the political perspectives across various topics. By using a clustering approach, we can experiment with different numbers of clusters and determine how many groups of opinions are reasonable in each case. This has the great advantage that we do not force the articles into a predefined and rigid model. This can be very helpful. For example if we assume that opinions on different topics can be classified into some major political tendencies, such as left and right or left, right, liberal and green, it might happen that for some topics perspectives overlap between these categories. Furthermore, clustering offers a wide range of approaches, so that we have the possibility to experiment in order to achieve the best result.

In clustering, data items are represented as points in a space. They are vectors of a certain dimensionality. Their relation is defined through distance metrics, for example Euclidean distance. However, the Euclidean distance metric does not represent every kind of data. Other distances, such as Jaccard or cosine distance are further options to compare data points.

On a general level, there are two main groups of clustering approaches: Hierarchical clustering and point-assignment. In hierarchical clustering, particularly agglomerative clustering algorithms, each point represents its own cluster in the beginning. Then, clusters are merged up to a certain point. This point depends on the stopping criterion, which might be the number of clusters, a certain dis-

tance threshold or the measure of compactness [65]. For the last case, as soon as compactness values for each cluster impair, merging stops.
The second group of clustering algorithms uses point-assignment. K-Means clustering is the most popular representative of point assignment clustering approaches. The minimum input into the K-Means clustering algorithm usually is "k", the number of clusters and a way of initialization, that chooses k centroids. Then, in a loop, all points are assigned to their closest cluster. After each iteration the position of the centroid is adjusted. The algorithm stops either after a certain number of iterations or if all points stay in the same cluster throughout multiple iterations[51]. These two approaches differ in their performance. While K-Means as a representative of the point assignment group has linear complexity, hierarchical clustering is quadratic in the number of documents[63]. In our case, we do not need to cluster a lot of documents. Therefore, we are not limited to either of those approaches due to performance issues. As this is an experimental setting we evaluate one representative algorithm for each clustering approach. By using K-Means and hierarchical clustering, we can also experiment with different distance metrics. K-Means is based on Euclidean distance, while hierarchical clustering works with various distance metrics. As some preliminary clustering experiments show very similar results for K-Means and hierarchical clustering with the use of Ward as a distance metric, which is based on Euclidean space, we use cosine and Manhattan distance for the hierarchical clustering. Cosine distance also corresponds to the approach of [63].
There are other clustering approaches, such as density-based clustering or fuzzy clustering [65] that we do not consider within the scope of this work.

6.3 Filtering Component

The last step of a content-based recommender system is the so-called "filtering component". In a traditional recommender system this component matches the user profile to the items and recommends items accordingly. In our case, we do not have user profiles. We rather have to choose one news article per political perspective to show to the user and thereby cover a broad political spectrum. To achieve this, we have to choose representative items out of each cluster of a political perspective, to recommend to the user. The most intuitive way to do this is to choose the item that is closest to the centroid of the cluster as a representative of that cluster. Therefore, upon completion of the clustering, we iterate through the clusters in order to choose one representative article for each cluster. In K-Means clustering, we choose the article whose Euclidean distance to the centroid is minimal. For the agglomerative clustering, we choose the article whose distance to all other points in the same cluster is minimal. By recommending one item out of each cluster, we should cover the various political perspectives related to the specific topic. Our data contains articles from eight different political parties. However, we do not expect that these political parties have clearly separable opinions on topics. Therefore, we experiment with different numbers of clusters to find the appropriate number. Regarding a possible resulting application, we expect people to read between three and five articles per topic, as time is a very limited resource. This implies that further filtering may be necessary in a later step in case the best solution has more than five clusters.

Code	**Label**	**Code**	**Label**
per000	undefined	per410	productivity +
per101	foreign special +	per411	infrastructure +
per102	foreign special -	per412	controlled economy +
per103	anti-imperialism +	per413	nationalization +
per104	military +	per414	economic orthodoxy +
per105	military -	per415	marxist analysis +
per106	peace +	per416	anti-growth economy +
per107	internationalism +	per501	environmentalism +
per108	europe +	per502	culture +
per109	internationalism -	per503	social justice +
per110	europe -	per504	welfare +
per201	freedom/human rights +	per505	welfare -
per202	democracy +	per506	education +
per203	constitution +	per507	education -
per204	constitution -	per601	national way of life +
per301	decentralism +	per602	national way of life -
per302	centralism +	per603	traditional morality +
per303	gov-admin efficiency +	per604	traditional morality -
per304	political corruption -	per605	law and order +
per305	political authority +	per606	social harmony +
per401	free enterprise +	per607	multiculturalism +
per402	incentives +	per608	multiculturalism -
per403	market regulation +	per701	labour +
per404	economic planning +	per702	labour -
per405	corporatism +	per703	agriculture +
per406	protectionism +	per704	middle class +
per407	protectionism -	per705	minority groups +
per408	economic goals	per706	non economic groups +
per409	keynesian demand +		

Table 1: Manifesto Codes and Labels.

6.4 Validation

The next step consists of the adjustment of parameters for the agglomerative clustering as well as the validation of the model. Linkage criterion and distance metric are the parameters to adjust in the agglomerative clustering algorithm. We use the following validation measures for both, the parameter adjustment and the final evaluation of the results. There are different methods to validate clustering approaches, known as *cluster validity*. On a general level, the meth-

ods divide into two groups, depending on the criteria they use for validation: *external* and *internal* cluster validation methods[65].
External measures take into account prior knowledge on the data. For example, if there are known class labels for some parts of the data, it is possible to evaluate how accurately the clustering algorithm assigned samples to their corresponding class. As we take the articles from political party related twitter accounts, we have the corresponding political party as a label for each article. This enables us to use these external cluster validation measures, such as completeness, homogeneity and the v-measure. The Rand index is another option for an external measure. We compare different clustering validation measures to see, which of them return the most appropriate results for our clustering problem.
Internal measures solely focus on the data and analyze the resulting clusters. In case that there is no ground truth, internal measures represent a good way to determine the appropriate number of clusters. More specifically, they analyze how *compact* clusters are through intra-cluster variance and how well they are *separated* from each other through the distance between their centroids. As these values have opposing trends there are more complex measures that capture both in one value. We use the silhouette coefficient [54]. Another possible measure to do this is the *Dunn-Index* [65].

- **Silhouette Coefficient:** The silhouette coefficient is a coefficient based on intra- as well as inter-cluster distances. It is defined as follows for each sample i [54]:
$$s(i) = \frac{(b-a)}{max(a,b)} \tag{5}$$
with a as the mean intra-cluster distance and b the distance between a sample and the nearest cluster that the sample is not part of. We calculate the mean coefficient over all samples. The result range of the silhouette value is: $-1 \leq s(i) \leq 1$. High values imply that a sample is near to the samples in its own cluster and far to the samples in the other clusters. 0 indicates that the distance of a sample to other clusters is the same as the distance of a sample to its own cluster and negative values mean that the sample is actually closer to other clusters than to its own. Various distance measures are applicable to the silhouette coefficient. We apply Euclidean distance.

As external cluster measures, we use the following:

- **Completeness [53]:** This measure indicates, whether all data points of a class are located in the same cluster. The result range of this measure is between 0 and 1, where 1 means that all data points of a class are in the same cluster.

- **Homogeneity [53]:** This measure indicates, whether all points within a created cluster belong to the same class. The result range of this measure is between 0 and 1, where 1 means that a cluster only contains samples of the same class. Completeness and Homogeneity have the following relationship: $homogeneity(a, b) == completeness(b, a)$.

- **V-measure [53]:** This measure represents the harmonic mean between the previous two measures, completeness and homogeneity.

$$v = 2 * \frac{(homogeneity * completeness)}{(homogeneity + completeness)} \tag{6}$$

  Accordingly, its values also lie between 0 and 1, with 1 indicating an optimal partition of the data.

- **Adjusted Rand Index:** The raw Rand index measures clustering quality by comparing pairs of points in two clusterings. Its values depend on whether the pairs of points are in the same or in different clusters among the compared clustering approaches. In its adjusted form, the adjusted Rand index is adjusted for chance and thereby accounts to the fact that we do not have a lot of samples per topic [24].

$$RI = \frac{A}{\binom{n}{2}} \tag{7}$$

  with A representing the number of agreements between the result set and the reference label set.

$$ARI = \frac{(RI - Expected\_RI)}{(max(RI) - Expected\_RI)} \tag{8}$$

  Values range between -1 and 1, where a score of 0 stands for random labeling and a score of 1 for a perfect match. Negative values appear if the actual index is smaller than the expected index. Hubert and Arabie [24] provide further detail on the chance adjustment.

Apart from cluster validity, we collect further statistics on the data, such as the number of samples, the class variation within a data set and the average length of an article. These values potentially influence the quality of the clustering results.
While these cluster validation approaches serve as quantitative evaluation instruments, it is as important to qualitatively analyze the results. As the main objective is to bring benefit into news consumption by opening up perspectives on topics, we have to prove if the recommended articles fulfill this goal and do not just duplicate content. Hence, the final recommendation set runs through a manual evaluation step in order to estimate if the recommended articles bring in new input, but also how well they cover the political spectrum. The definition of this spectrum plays an important role, which we shall address in the following section.

6.4.1 Defining Political Perspectives

We already discussed the difficulty of a clear definition of political perspectives in section 5.5. Nevertheless, we have to provide reference labels in order to be able to apply the external validation measures. We use political parties. While we could group political parties into left and right, it is questionable, whether the very high-level “left-right” scale is an appropriate representation for different political perspectives across topics [45]. Certainly, there are topics that are clearly related to the left- or the right wing of politics, but it is hard to imagine that every topic fits into these categories. Sometimes, we might even find similar positions in the extreme left and the extreme right. For example, if we consider the refugee question in Germany, where both extreme sides appear more restrictive than the political center. Looking at the political landscape in Germany, apart from left and right, we additionally have the Green party and the liberal party, which in many cases would be hard to include into a two-dimensional “left-right” scale However, there are attempts to position all parties on a left-right spectrum. The seating order of the German Bundestag roughly represents this two-dimensional political spectrum following the example of the first French national assembly during the French Revolution [69]. We consider eight political parties as reference labels that based on this criterion follow this order:

Figure 5: Political Spectrum in Germany

This is not a comprehensive list of political parties that act in Germany but the choice represents the whole spectrum. While we use the individual parties as references for external validation, this spectrum helps to determine how well the recommendations cover political directions in the qualitative validation of recommendations. We refer to it as the traditional political spectrum from now on.

7 Implementation

In this section we define the technical realization of the approach we describe in previous section 6. We realize our approach as a Python program. We run the Python program on a virtual machine running Ubuntu 15.10 (64 bit) with 1.8 GB RAM, 64 MB graphics memory and 16 GB disk space on a SSD. The Python version is 2.7.10. We divide the approach into four major steps: Content Analysis, Clustering, Validation and Recommendation. From an initial pool of articles they choose two to five articles, each of which represent a different political viewpoint. Figure 6 illustrates the high-level architecture of the recommender.

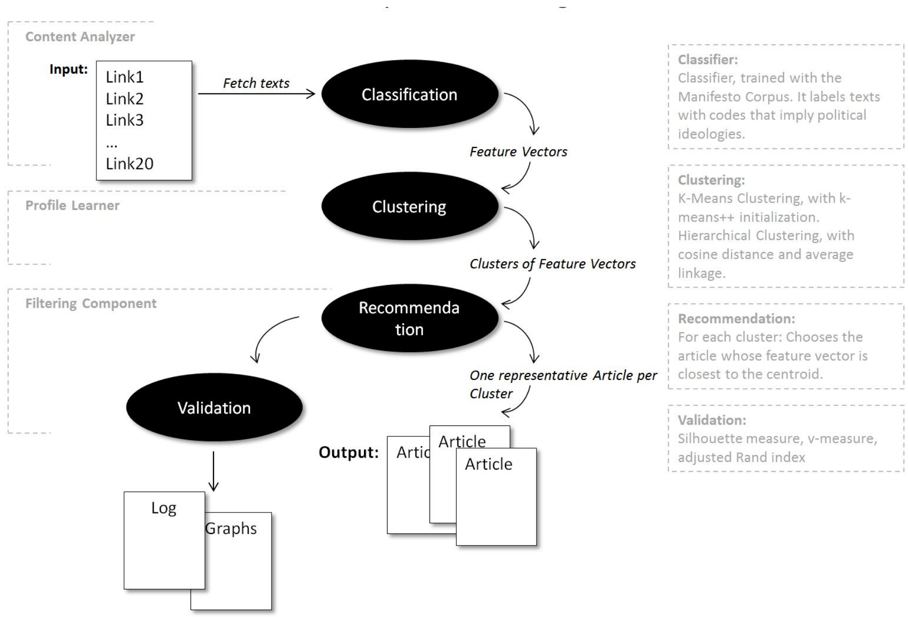


Figure 6: High-level Recommender Architecture

7.1 Content Analyzer

The goal of the content analyzer is to create three feature vectors for each article of a topic:

1. The Manifesto label vector
2. The BOW vector
3. The TF-IDF vector

As an input, we provide a list of links in a text file. They represent a collection of articles on a restricted topic, such as the refugee crisis, or the financial crisis of Greece. The topic restriction comes with the input data. In a first step, we extract the plain article text from these links and save it to text files, as an input for the feature extraction. Therefore, we use the *fetch_url* method that the Manifesto classifier provides. The code of the Manifesto classifier is available on GitHub [7]. It retrieves the article title and text from a given URL, with the help of the *BeautifulSoup* package that helps to extract only relevant text from an HTML file.
To extract the Manifesto labels, we then train the classifier on the Manifesto dataset before we use it to predict Manifesto label values for all articles. The training and classification pipeline contains the following components: *CountVectorizer*, *TfidfTransformer* and *LogisticRegression*. With the

[7] https://github.com/felixbiessmann/fipi

GridSearchCV method the approach tests various parameter settings for the three components and chooses the best result. The output is a normalized 57-dimensional vector for each article. This differs from the original version of the classifier as it leaves out the mapping of the individual prediction values to a left-right scale as well as to a smaller number of domains.
The *sklearn* package provides the required functionality for the extraction of the BOW and TF-IDF vectors. With the *CountVectorizer* and the *TfidfVectorizer*, we extract the two vectors, while removing stop words. We normalize the vectors with the L1 norm.
As an output of this component, we store three data files that contain the feature vectors of the articles. We use these as an input to the clustering component.

7.2 Clustering

In the clustering, we aim to group articles together that belong to the same political perspective based on the three extracted feature sets. For this purpose we use the *sklearn* library for python that provides a wide range of machine learning algorithms. We follow two clustering approaches: K-Means and hierarchical clustering. We sequentially perform the two clusterings on the three data sets. The input to both clusterings are the feature vectors. We load the corresponding files and call the clustering method. We use the *K-Means* and the *AgglomerativeClustering* methods from the *sklearn.cluster* package. The input to the K-Means clustering is the data, the number of clusters and the initialization method. We run the clusterings with two to eight clusters and use the default number of ten initializations on different centroid seeds as well as the default k-means++ initialization method in a first step. In contrast to random initialization, it spreads the initial centroids across the space. In this case, we set a seed in order to generate reproducible results. In a later step, we apply a more sophisticated initialization described in 7.2.1. The K-Means clustering returns a vector with cluster assignments, as well as the resulting centroids of the clusters. As a second clustering approach we choose agglomerative clustering as a representative method for hierarchical clustering. Apart from the data, the method requires distance measure and linkage criterion as input parameters. We experiment with cosine and Manhattan distances as well as with average and complete linkage. We dedicate the experiment described in 9.2 to the adjustment of these input parameters. The agglomerative clustering also returns a vector of cluster assignments.

7.2.1 Initialization

We apply a second initialization approach for the K-Means clustering on Manifesto labels. Therefore, we run the Manifesto classifier on the current political manifestos of the political parties in question and use their resulting normalized label vectors as initial centroids for K-Means clustering.

7.3 Cluster documentation

In the cluster documentation we process the cluster assignments of both clustering approaches in order to figure out which cluster approach and which number of clusters returns the best results. We use common cluster metrics to evaluate cluster quality. Our method *cluster_documentation*, takes the data and the clustering object, which includes cluster assignments and centroids and the target labels as an input. The *sklearn.metrics* package provides the following methods for the measures:

- silhouette_score: cluster assignments, distance metric
- homogeneity_score: cluster assignments, target labels
- v_measure: cluster: assignments, target labels
- adjusted_rand_score: assignments, target labels

We calculate the silhouette coefficient with Euclidean distance as a metric. In addition to these measures, we calculate the inner cluster sum of squares SSE and store individual silhouette graphs for each cluster. We record these results for both clustering approaches and all numbers of clusters in a log file for later evaluation.

7.4 Recommender

The recommender represents the filtering step of a traditional recommender system. Its objective is to define which articles to recommend out of the clustered data. In our implementation, the method *find_representative_articles* fulfills this task. It takes the data and the cluster assignments of the articles as an input. In this stage of the recommender design, we choose to recommend one article per cluster. In a later application this number may have to be reduced. For K-Means clustering, the centroids are already given in the clustering object. We iterate through all the clusters. Then, for all articles in a cluster, we calculate the Euclidean distance to the cluster centroid and choose the article that is closest to it. For the agglomerative clustering there are no defined centroids. Therefore, we calculate pairwise distances between all points in each cluster. As we use cosine distance to merge the clusters in the actual clustering step, we also use cosine distance to quantify the distance between points in this step. We then choose the point of the cluster that has the minimum cosine distance to all other points in the same cluster as a representative. The method returns a list of article IDs, one per cluster.

8 Data

We experiment on two different data sets of news articles on recent political topics: A very small data set for the evaluation of the Manifesto classifier and a larger data set of articles that have been published in political twitter accounts.

We manually select the articles for various reasons. First of all, to our knowledge, there is no data set publicly available that contains the information we require. For validation, we need an assignment of political perspectives for the articles. This could be a political party or a political direction. Secondly, we aim to keep the amount of data limited. Even though there are several ways to automatically evaluate cluster quality, it is important to comprehend the results based on the actual data and be able to actually read it in order to assess the possible benefit of the recommender system.
For each article we store its source link, as this is the input to the classifier, the topic the article belongs to as well as a political label. For the Twitter articles, we additionally keep the related tweet. Most clustering approaches are sensitive to the sizes of their clusters. Therefore, it is advantageous to have approximately equally sized clusters. However, political parties focus on certain political topics to sharpen their profile. They particularly promote topics that benefit their electoral campaign. Considering the Twitter profiles of political representatives of political parties, or the party profiles themselves, we find a very unequal distribution of topic per profile. As an example a critique on a political decision of a party that is politically in charge is a great opportunity for distinction of other political players. Certain political parties are also traditionally bound to certain topics more than others. For example the Green party frequently comments on developments in nuclear energy, while other parties do not mention this topic. Despite this different representations, we aim to equalize the number of articles of a certain topic per political category as it improves the clustering potential, but also to bring up a maximum of different opinions on each topic.

8.1 Data Toy Example

As a starting point we construct a toy example, that evaluates the functionality of the classifier and represents a basis for the further implementation. We choose two topics for which we manually search for a small number of news articles. We particularly look for articles that represent different political opinions. To do this, we diversify the publishing news papers, as some news papers are known to represent rather left or right opinions. As we set the topics in advance, we use Google News to find relevant articles for the preselected topics. These topics are very controversially discussed political topics from the past year. For the articles of this data set, we collect the source and assign a manual label that classifies the data into left, neutral or right.

- **Referendum Greece, 2015:** The referendum of 2015 in Greece, in which Greek people are asked to vote for or against European austerity packages. The referendum was questioned a lot at that time as people considered it to really be about the choice, whether or not Greece should remain in the European Union. Particularly from Germany, there were strong opinions on this question as Germany bears a big proportion of the financial aids going to Greece. German people were concerned to be investing their tax payments into a country, which will not follow the given agreements.

- **Upper Limit for Migration:** The second topic is a discussed upper limit for the number of refugees migrating to Germany, in the beginning of 2016. A lot of of discussions raised, when Austria announced to introduce such an upper limit. Afterwards, there were discussions on the decision of Austria, as well as the introduction of an upper limit for Germany as well.

The choice of these two topics is due to the expectation of extreme reactions in the news.

8.2 Twitter Data

For the main experiments, we collect a second data set of ~ 200 news articles. We collect these articles from Twitter accounts of eight German political parties and their representative politicians. The articles either have been tweeted or retweeted by the corresponding account. As political parties use Twitter to promote their activities and opinions, we believe these articles to be a good representation of their perspectives. They will rather express different opinions than only covering the facts. We choose articles on currently discussed news topics. We shortly describe each of the topics whose clustering we evaluate in more detail in order to provide insights and understanding for the outcomes.

- **Refugee Crisis:** As a result of the war in Syria more than a million migrants came to Europe in 2015, which confronted European countries with operative and administrative challenges. Germany stood out in this discussion, as it reacted with a welcoming attitude.

- **Right-wing Populism:** Also, as a result of the refugee crisis, but beginning with the financial crisis of Greece, more radical parties came to more power in German parliament. Especially, a newly founded party called "AFD" that was created as a response to the Greek crisis and strongly voted for an exclusion of Greece of the European Union.

- **Federal President Gauck:** The president announced in June 2015 that he would not prolong his candidature in the upcoming year. The majority of people as well as politicians were supporting him and were sorry for his decision.

- **AFD Politician Gauland:** Gauland is a politician that raised attention because of a comment on the German national soccer player Boateng by saying that people want him to play soccer but would not want to have him as a neighbour. While he tried to argue that he did not know that the player was born in Germany and had the German citizenship for his whole life, politicians from all other parties stood on Boateng's side and condemned racism.

- **Safe Countries of Origin:** As asylum applications increase with the refugee crisis. There are discussions about declaring Northern African countries to safe countries, which will only allow asylum applications in case of political refugees. Especially, the Green party was opposed to that change.

- **Big party crisis:** The two strongest political parties in Germany: CDU and SPD have lost a very big proportion of their power. This also enabled more extreme parties to gain power and often is referred to as a threat for democracy. Sources for this development are discussed very controversially.

- **Turkey - Erdogan:** For multiple reasons the relation between the European Union and Turkey is tense. Firstly, because of the refugee deal that has been made and secondly because of Turkey's treatment of press freedom. Multiple journalists and politicians have been warned or even sued because of expressing their opinion on the political situation in Turkey.

- **Brexit:** On June 23rd the United Kingdom had a referendum on leaving the European Union. In the end 51.9% of people voted for leaving the European Union, while only 48.1% of the people voted for remaining in the European Union. Even though predictions on the results showed that it would be very close, a lot of people were surprised about the outcome, including British people themselves. The referendum was and still is a very hot topic in the press, as consequences are not fully known. As a next step, relations in terms of trade and migration between the UK and the European Union have to be negotiated. Furthermore, London's role as financial capital of Europe is in question. UK's prime minister David Cameron stepped down and no potential successor has been defined.

While some of these topics are very controversially discussed, others are rather homogeneous. For example, the refugee crisis currently is a source of political conflict between parties in power. On the other hand president Gauck's statement on not prolonging his candidature did not raise as contradictory voices. The question on how narrow to define a topic is a conceptual choice for the recommender system. News are very dynamic and especially political news are usually based upon some events, statements or decisions. However, a sequence of these events, statements or decisions will belong to the same topic as they often are reactions on previous actions. For the recommender this means that there is the possibility to choose high-level topics that comprise a longer time period, which can be weeks or months and therefore a sequence of actions. On the other hand, a very low-level definition would consider each action individually. As an example, the referendum for the Brexit in June 2015 is a high-level topic. Previous predictions on it, direct reactions on the outcome, the following resignation of politicians are all individual actions and low-level topics that raise discussions in the news. A recommender that aims to deepen and improve political knowledge can work on both, high-level and low-level topics. For a high-level topic, the recommender brings in different aspects on a topic, while for low-level topics, it shows opinions and perspectives on single actions. The high-level topic recommendations carry the risk of actually displaying different subtopics on a political question rather than view points on the same action. However, this reduces the risk of displaying duplicate content to the reader, which seems inevitable for a plurality of articles on a very narrow topic. For this reason, we choose rather high-level topics for our experimental setting. We consider all the above mentioned as such. Still, we consider different time spans, from months for the refugee crisis, over weeks for the Gauland-Boateng discussions, to four to five days for the Brexit.

8.3 K-Means Initialization Data

We collect a further data set for the initialization procedure described in 7.2.1. It comprises current manifestos of all political parties that we consider in the articles of our data set. These programs have between 43 and 190 pages in total. They provide detailed information on the parties' fundamental political principles by describing their position in various topics, such as democracy, education or the social market economy. While they are similar to the manifestos on which the classifier was trained, these are electoral programs of the political parties, directed to a certain election. The nature of the texts differs from a news article. However, they cover many dimensions topic- and perspective wise. Therefore, we expect them to be a good representation of the general political direction of an political party and a good centroid, respectively.

9 Experiments

We conduct the following experiments to find out whether the Manifesto classifier is an appropriate instrument for political text characterization and which clustering approach best applies to the dimensional space of Manifesto label vectors. We present a sequence of experiments that build on each other and explore and analyze the data as well as the problem itself. The results of one experiment define the set up and validation of the following. We first evaluate the clusterings per topic, before we summarize and draw general conclusions on the results. As every topic is different in terms of external perception and controversy in news coverage, this enables us to estimate the impact of the actual data, by evaluating individual articles, but also to see if the results are generalizable.
First, we explore the behavior of the classifier on the small toy data set. Based on the outcomes of this analysis, we design more complex experiments. They involve more data and a greater variety of validation measures. Before the actual cluster analysis, we compare different parameter settings in section 9.2. With the most suitable parameter set, we start the cluster analysis on the Twitter data. Due to different amounts of articles per topic, we divide the experiment into three parts: We start with articles on the Brexit as this topic has a reasonable number of articles from a rather short period of time. Then, we evaluate the refugee crisis and right-wing populism as two more topics with around twenty

articles each, but longer time spans, in order to validate if the findings of the Brexit are consistent in further topics. We apply two clustering approaches on each topic: K-Means and hierarchical agglomerative clustering and we have three different sets of features per topic. In a last step, we prove whether a smaller number of articles impacts the result with an analysis of the remaining topics with ten to twenty articles. Depending on the number of articles we vary the maximum number of clusters. Table 2 visualizes this experiment setting. In

Topics with > 20 articles			
Topic	**Clusters**	**Approach**	**Feature Set**
Brexit	2-6	Hierarchical	BOW
Refugee Crisis	2-7	K-Means	Manifesto Labels
Right-Wing Populism	2-7		TF-IDF
Topics with ≤ 20 articles			
Topic	**Clusters**	**Approach**	**Feature Set**
Gauck	2-5	Hierarchical	BOW
Gauland	2-5	K-Means	Manifesto Labels
Safe Countries of Origin	2-5		TF-IDF
Party Crisis	2-5		
Turkey	2-5		

Table 2: Experiment setting

a last step, we conduct some side experiments that do not directly tackle our research question but provide better knowledge of the data. In particular, we perform a cross-topic clustering to see, whether it is possible to identify political parties by their language on a cross-topic basis.

9.1 Left-Right Classification

We first experiment on the two topics of our data toy example: the Greek referendum on remaining in the European Union (Topic 1) and the upper limit for migration to Germany as part of the refugee discussion (Topic 2). We conduct this experiment to get familiar with the classifier. Furthermore, we aim to see, whether the classification into left and right corresponds to our own assessment of the data.
The classifier in its basic version provides us with three values on a given text or link:

- **Left-Right Value:** Value between 0 (left) and 1 (right)
- **Domain:** External Relations, Fabric of Society, Welfare and Quality of Life, Freedom and Democracy, Political System, Economy
- **Label:** Military, Peace, Political authority, ... (full list in Table 1)

First, we get a value for a left-right scale between 0 and 1. In practice, an article is always assigned to a political direction, left or right with a certain percentage whenever it crosses the 0.5 border. We transfer this value to a scale between 0 and 1, where 0 represents left and 1 represents right. To achieve this, we subtract all results that have been classified as left from 1. In addition to the left/right classification, the original classifier returns the domain of an article and the Manifesto labels it has been assigned to. The classifier displays the domain with the highest fit and a corresponding value, between 0 and 1.
For each topic, we have a set of 20 related articles. As a validation in this first step, we manually label these articles with the following labels: *left*, *neutral* or *right*. *Left* corresponds to anti-capitalistic perspectives, *neutral* stands for a mostly objective way of reporting and *right* represents rather conservative views.

9.1.1 Results

Independent from the topic, the distribution of the left-right values of the observations is remarkable, as values accumulate in the extreme left or the extreme right. There are no values in the middle between 0.4 and 0.6, which would represent a neutral position of the news report (Figure7).

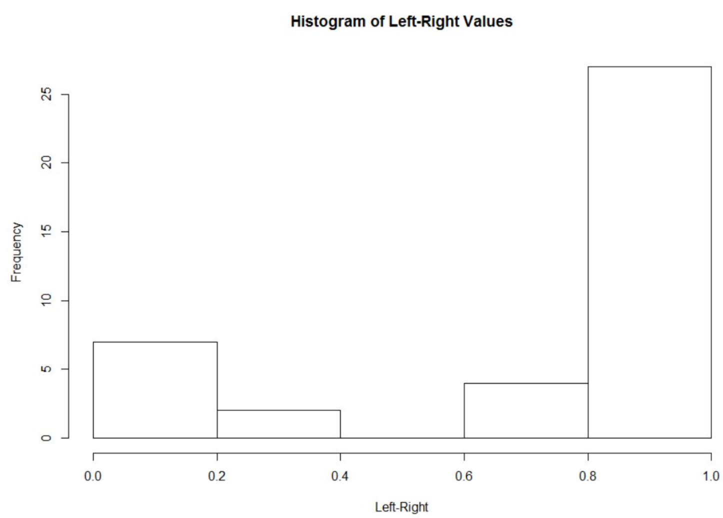


Figure 7: Histogram of Value Distribution

Topic 1 has a variance of $v_1 = 0.1489$, while topic 2 shows lower variance with $v_2 = 0.0866$. Figure 7 further shows that there is a clear trend to the right. The corresponding arithmetic mean values are $m_1 = 0.6635$ and $m_2 = 0.841$.
Tables 3 and 4 contain detailed values of this first experiment. We manually categorized most articles as neutral. Blanks in the manual labels represent articles, which we could not assign to either of the three categories. We find that our manual observations contradict the extreme left and right positions

ID	ML	LR	Domain	Label
1	neutral	1	Political System, 0.7	political authority, 0.7
2	left	**0.73**	External Relations, 0.76	europe, 0.76
3	neutral	0.72	Political System, 0.72	political authority, 0.67
4	neutral	0.98	Political System, 0.97	political authority, 0.97
5	neutral	0.96	External Relations, 0.55	europe, 0.54
6	left	**1**	Political System, 0.99	political authority, 0.99
7	-	0.89	External Relations, 0.99	europe, 0.99
8	neutral	0.02	External Relations, 0.88	internationalism, 0.47
9	neutral	0.92	External Relations, 0.52	europe, 0.5
10	left	**0.9**	External Relations, 0.85	europe, 0.85
11	left	**0.9**	External Relations, 0.85	political authority, 0.76
12	neutral	0.72	Freedom and Democracy, 0.57	freedom/human rights, 0.57
13	neutral	0.01	External Relations, 0.78	internationalism, 0.53
14	-	0.98	External Relations, 0.98	europe, 0.98
15	neutral	0.08	External Relations, 0.95	europe, 0.95
16	neutral	0.3	External Relations, 0.67	europe, 0.64
17	neutral	0.11	External Relations, 0.8	europe, 0.79
18	left	**1**	Political System, 0.97	political authority
19	left	0.05	Economy, 0.46	anti-growth economy
20	right	1	Political System, 0.77	political authority, 0.62

Table 3: Result Table for the Referendum in Greece
ID = ArticleID, ML = Manual Label, LR = Left-Right-Value

that the classifier returned.
We assume a classification of a neutral article to any of the political directions to be less severe than a classification from a left or right article to the opposite political direction. If we consider only the articles to which we manually assigned a political direction and that we did not label with neutral, we have seven articles from topic 1 and nine articles from topic 2. In that case, for topic 1, 71.43% of the articles are mismatched in comparison to the classifier results, meaning they are classified in the opposite political direction. For topic 2, the mismatch amounts to 44.44%. The consideration of domains and labels shows that they roughly correspond to what one would expect for the two topics. Political authority and Europe very well describe dimensions of the financial crisis and freedom/human rights also represent a main aspect of the refugee discussion. However, the domain refers to the topic and not to the actual perspective. Likewise, a single label is insufficient for text characterization.

9.1.2 Interpretation

The process of the logistic regression classification, in which values are projected to a sigmoid-shaped curve, favours values close to the extremes zero and one. This may explain the absence of neutrally classified articles.

ID	**ML**	**LR**	**Domain**	**Label**
21	neutral	0.99	External Relations, 0.74	europe, 0.72
22	neutral	0.04	Economy, 0.58	market regulation, 0.54
23	right	1	Political System, 0.99	political authority, 0.99
24	left	**1**	Political System, 0.88	political authority, 0.87
25	neutral, right	0.93	Freedom and Democracy, 0.45	freedom/human rights, 0.41
26	neutral, right	1	Freedom and Democracy, 0.93	freedom/human rights, 0.66
27	right	1	Freedom and Democracy, 0.74	freedom/human rights, 0.68
28	left	**0.99**	Freedom and Democracy, 0.79	freedom/human rights, 0.79
29	neutral	0.71	External Relations, 0.87	europe, 0.59
30	neutral	1	Freedom and Democracy, 0.61	freedom/human rights, 0.57
31	neutral	1	Political System, 0.98	political authority, 0.97
32	neutral	0.97	External Relations, 0.98	europe, 0.77
33	neutral	0.99	Welfare and Quality of Life, 0.98	social justice, 0.87
34	neutral	0.94	External Relations, 0.59	europe, 0.57
35	-	1	Political System, 0.92	political authority, 0.9
36	left	**0.99**	Freedom and Democracy, 0.89	freedom/human rights, 0.67
37	neutral	0.03	External Relations, 0.83	internationalism, 0.42
38	neutral	1	Freedom and Democracy, 0.94	freedom/human rights, 0.94
39	right	0.99	Political System, 0.94	minority groups, 0.96
40	right	**0.25**	External Relations, 0.95	europe, 0.81

Table 4: Result Table for the Migration Limit in Germany
ID = ArticleID, ML = Manual Label, LR = Left-Right-Value

Another observation refers to the process of manual labeling: We found that some articles are very hard to categorize in terms of left and right. A reason for this is that they discuss left and right opinions within the same article without taking sides. One article even questions the meaning of left and right in the context of the refugee discussion, more particularly Merkel's decision on keeping the borders open and concludes by stating that this is neither right nor left it is simply moving on [8].
Regarding the general trend to the right that the Manifesto classification shows, the nature of the topics can explain that the Greek Referendum has more balance between left and right than the upper limit for migration. The discussion on the referendum in Greece and the whole financial crisis is a capitalistic topic. The question on the role of banks, debts and loans is omnipresent. More importantly, it is one, where the positions of the political left and right are rather clear. The political left criticizes the power of the banks and supports the left government of Greece that organized the referendum. They want to protect the people of Greece from cuts in social welfare areas. The political right supports the "Grexit", the exclusion from the European Union instead of financial support for the country, while the political center is trying to negotiate and set up regulations and austerity plans that will bring the country back to the financial market with their financial support. On the other hand, in the refugee discussion positions cannot be clearly related to one political side. In general, one would assume that the political right is in favour of closing up the boarders, while the political left would follow an approach of open boarders. However, the government officially communicated the willingness to welcome refugees, which led to criticism, from both sides, left and right. And as the topic of the articles refers to closing up boarders in order to reduce migration, it is comprehensible that an automated text analysis can relate these texts easier to politically right positions than to the opposite.
Even though the difference in the nature of the topics may explain the general trends of the average values of the classification, the mismatch of manual and classifier labels is way to high to build a trustworthy model on this basis. As the manual categorization already showed, it might not be appropriate to categorize political perspectives on two dimensions only. A deeper investigation on the articles also shows that individual parts of the text have very contradicting results. A single value for the whole text can therefore easily lead to inappropriate results even though it contains clearly left or right parts. This can also be a result of the classification approach: The left and right classification is based on a logistic regression model that actually assigns probability values to 57 different labels representing political ideologies. The final left and right assignment is the result of the summary of those labels into two dimensions, by simply mapping certain labels to left and others to right. Out of 57 labels, the classifier uses 26 labels to map to left and right. Table 5 shows the labels used for the mapping. This mapping is hard coded and not empirically validated.

[8] http://www.sueddeutsche.de/politik/fluechtlingspolitik-linke-runter-von-der-wohlfuehlinsel-1.2952704-2

Right		**Left**	
104	military +	103	anti-imperialism +
201	freedom/human rights +	105	military -
203	constitution +	106	peace +
305	political authority +	107	internationalism +
401	free enterprise +	403	market regulation
402	incentives +	404	economic planning +
407	protectionism +	406	protectionism +
414	economic orthodoxy +	412	controlled economy +
505	welfare -	413	nationalization +
601	national way of life +	504	welfare +
603	traditional morality +	506	education +
605	law and order +	701	labour +
606	social harmony +	202	democracy +

Table 5: Mapping from Manifesto Labels to Left and Right Categories

This is possible as these ideologies traditionally embody left or right political perspectives. Still, it simplifies the "original" label vectors and thereby carries a risk of information loss. To avoid this risk and for better representation, we from now on experiment with the whole 57-dimensional vector of ideological labels.

9.2 Parameter Adjustment

There are two types of parameters to adjust in the agglomerative clustering implementation: The linkage criterion and the distance metric. We apply the clustering on the news articles of the Brexit topic, to find the best values for these parameters. First, we compare *average linkage* to *complete linkage*, keeping the distance metric stable with the cosine distance metric. The linkage parameter defines, which distance the algorithm uses to relate sets of observations to each other. Average distance uses the average distance of samples between two sets of observations, while complete linkage uses the maximum distance between samples in two sets of observations as a minimization criterion, when merging two clusters [59]. The second parameter defines the distance metric used to calculate these distances. We experiment with cosine and Manhattan distance.

9.2.1 Results

Table 6 shows silhouette values, v-measure and ARI for all three feature sets and opposes average linkage to complete linkage, respectively. In the last column it displays "Yes" in case the change of linkage type influenced the resulting recommendations and "No" if it had no effect. As it does influence the results in more than 50% of the cases, we further evaluate the effect. For the BOW-features, the results are equal for two to four clusters. For five and six clusters, average

linkage brings better silhouette values as well as v-measure and the ARI.
The Manifesto labels show better results overall and again average linkage outperforms complete linkage for all three measures. For the TF-IDF feature set, results only differ for five to six clusters, mostly in favour of complete linkage. We will use average linkage in our further evaluations for two reasons. In order to keep the feature sets comparable, parameters should be equal for all three feature sets. As the Manifesto labels are the main subject of our experiments and the other two sets act as baselines, we choose to reach the possibly best values for the Manifesto labels.
From now on, we keep the average linkage as a comparison criterion between observation sets stable, while we experiment with Manhattan and cosine distance. We do not use Euclidean distance, as K-Means clustering already uses Euclidean distance. Table 7 shows that the choice of distance only influences around 25% of the recommendations. For the BOW-labels the cosine distance brings worse results than Manhattan distance for the silhouette coefficient. The v-measure and the ARI in average bring greater values for Manhattan distance than cosine distance. In contrast, cosine distance brings greater values for all measures with the Manifesto Labels, but not for the TF-IDF features. Here, cosine distance performs better for the silhouette coefficient and the ARI, but not for the v-measure. As a distance metric, we choose the parameter that optimizes the Manifesto Label set and further calculate with cosine distance.

9.3 Brexit

For the Brexit, we found 24 articles tweeted by six different politicians or political parties. Table 8 shows the amount of articles per party. The average length of an article for this topic is 630 words. Based on the results of the parameter adjustment, we use average linkage as a linkage criterion and cosine distance as a distance measure for the agglomerative clustering. We start with the K-Means++ initialization of the K-Means clustering. We partition the data into two to six clusters corresponding to the maximum amount of different parties.
T1: Brexit
T2: Refugee Crisis
T3: Right-Wing Populism
T4: Federal President Gauck
T5: AFD Politician Gauland
T6: People's Party crisis
T7: Safe States of Origin
T8: Turkey

9.3.1 Results

Sihouette Coefficient:
Regarding the silhouette coefficient, the BOW feature set reaches the highest individual value of $s = 0.5954$ with two clusters for K-Means as well as agglomerative clustering. On the average however, over all numbers of clusters, the

	Silhouette			V-Measure			Adj.Rand			Diff
k	**avg**		**cmpl**	**avg**		**cmpl**	**avg**		**cmpl**	
					BOW					
2	0.5954	=	0.5954	0.0919	=	0.0919	0.0082	=	0.0082	No
3	0.2098	=	0.2098	0.1456	=	0.1456	-0.0028	=	-0.0028	No
4	0.2018	=	0.2018	0.1989	=	0.1989	-0.0066	=	-0.0066	No
5	0.1972	>	0.0544	0.2549	>	0.2212	-0.0011	>	-0.0225	Yes
6	0.1028	>	0.0627	0.3188	>	0.2748	0.0162	>	-0.0097	No
					Manifesto Labels					
2	0.4908	>	0.3842	0.0677	<	0.1016	-0.0097	>	-0.0268	Yes
3	0.4296	>	0.4001	0.1826	>	0.1516	-0.0183	>	-0.0278	Yes
4	0.4443	>	0.4218	0.3139	>	0.2734	0.0129	>	-0.0055	Yes
5	0.2937	<	0.5235	0.3560	>	0.3009	0.0178	>	-0.0352	Yes
6	0.2506	<	0.5729	0.3487	<	0.3798	0.0063	>	-0.0122	Yes
					TF-IDF					
2	0.5821	=	0.5821	0.0919	=	0.0919	0.0082	=	0.0082	No
3	0.1993	=	0.1993	0.1456	=	0.1456	-0.0028	=	-0.0028	No
4	0.1964	=	0.1964	0.1989	=	0.1989	-0.0066	=	-0.0066	No
5	0.2034	>	0.0109	0.2549	<	0.2580	-0.0011	>	-0.0034	Yes
6	0.2142	>	0.0108	0.3027	<	0.3186	0.0039	>	0.0012	Yes

Table 6: Comparison of Average Linkage and Complete Linkage in Agglomerative Clustering

	Silhouette			V-Measure			Adj.Rand			Diff
k	**cos**		**man**	**cos**		**man**	**cos**		**man**	
					BOW					
2	0.5954	=	0.5954	0.0919	=	0.0919	0.0082	=	0.0082	No
3	0.2098	=	0.2098	0.1456	=	0.1456	-0.0028	=	-0.0028	No
4	0.2018	>	0.1588	0.1989	<	0.2144	-0.0066	<	-0.0025	No
5	0.1972	>	0.0877	0.2549	<	0.2822	-0.0011	<	0.0145	Yes
6	0.1028	>	0.0252	0.3188	>	0.3077	0.0162	>	-0.0071	Yes
					Manifesto Labels					
2	0.4908	=	0.4908	0.0677	=	0.0677	-0.0097	=	-0.0097	No
3	0.4296	=	0.4296	0.1826	=	0.1826	-0.0183	=	-0.0183	No
4	0.4443	=	0.4443	0.3139	=	0.3139	0.0129	=	0.0129	No
5	0.2937	>	0.2735	0.3560	>	0.3071	0.0178	>	0.0023	Yes
6	0.2506	=	0.2506	0.3487	=	0.3487	0.0063	=	0.0063	No
					TF-IDF					
2	0.5821	=	0.5821	0.0919	=	0.0919	0.0082	=	0.0082	No
3	0.1993	=	0.1993	0.1456	=	0.1456	-0.0028	=	-0.0028	No
4	0.1964	>	0.1500	0.1989	<	0.2144	-0.0066	<	-0.0025	No
5	0.2034	>	0.0782	0.2549	>	0.2551	-0.0011	=	-0.0099	No
6	0.2142	>	0.0902	0.3027	<	0.3053	0.0039	>	-0.0072	Yes

Table 7: Comparison of Cosine Distance and Manhattan Distance in Agglomerative Clustering

Political Party	T1	T2	T3	T4	T5	T6	T7	T8
AFD	5	4	3	1	1	1	-	-
CDU	2	5	5	2	6	2	4	1
CSU	-	2	4	-	-	1	-	1
DIE LINKE	4	4	3	2	-	3	2	2
FDP	5	3	6	2	1	2	-	2
GRÜNE	4	-	4	1	2	1	5	1
NPD	-	8	-	-	-	-	-	-
SPD	4	4	3	4	3	2	-	3
Total	24	30	28	12	13	12	11	10

Table 8: Political Party Distribution among Articles per Topic

Manifesto feature set outperforms all other approaches with an average silhouette coefficient of 0.4572 and a top value of 0.5720 for six clusters. TF-IDF and BOW follow in rank. Values greater than 0.5 speak for a structure in the data. We crossed this threshold for all three feature sets.

V-Measure and Adjusted Rand Index:
While the silhouette coefficient is an internal measure, it is also important to evaluate external validation measures with respect to the political parties that tweeted the articles. Figure 8 shows the results of the v-measure and the ARI for the same experiment setting. As a general trend, the v-measure improves for an increasing number of clusters. The Manifesto labels with agglomerative clustering by far outperform the BOW feature set and the TF-IDF feature set at five clusters with the maximum value of 0.3560. As the v-measure is the harmonic mean of completeness and homogeneity, we are able to comprehend its composition. Homogeneity values, in all cases are lower than the corresponding completeness values. This means that while most of the category members are in the same cluster, the clusters are still not homogeneous and contain members of other categories This is fine, as we also experiment with lower cluster numbers than political parties. Therefore, multiple parties can be assigned to a single cluster. However, one party should not be spread across various clusters. The evaluation of the ARI shows worse results than the v-measure. Overall, the values are very close to zero, which implies a random labeling. Furthermore, the results and general trends are not consistent to the v-measure. However, we similarly achieve the best value through agglomerative clustering with the Manifesto feature set with a value of 0.0178. On average agglomerative clustering with BOW features outperforms the other approaches. K-Means clustering brings worse results than agglomerative clustering for all approaches.

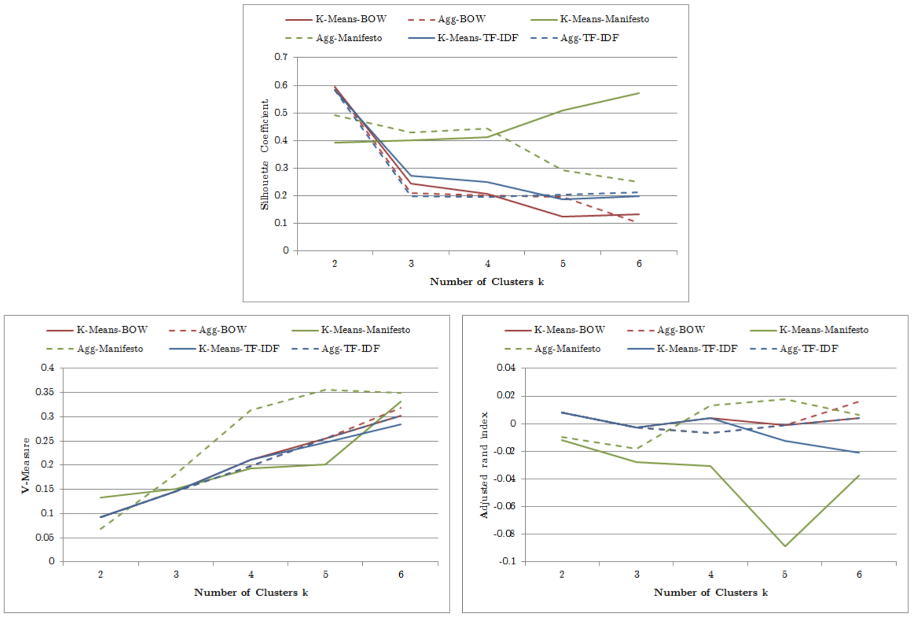


Figure 8: Brexit - Silhouette, V-Measure, ARI

Clustering:
For Agglomerative clustering, the biggest cluster contains 14 out of the 24 samples and there are two clusters that only contain one sample. For K-Means clustering, the biggest cluster still contains 10 samples and one cluster contains only one sample.

Recommendations:
In a last validation step, we evaluate the actual recommendations. The best performing combination in terms of v-measure and ARI: Agglomerative clustering on the Manifesto labels with five clusters, recommends articles that have been tweeted by the following political parties:

- **Aggl., Manifesto:** Grüne, DIE LINKE, FDP, AFD, SPD

This set of recommendations represents a very diverse spectrum of political parties with no re-occurrences of any political party. In order to ensure that this diversity of the tweeting political parties relates to actually diverse content of the articles, we read and compare the article texts as a last step of the Brexit analysis. From the readers perspective, each of the articles in the recommendation that comes from the Agglomerative Clustering with Manifesto labels, brings new information and little duplicate content. Table 9 summarizes this content and indicates information gain:

Origin	Content	Source	Inf.Gain
Grüne	short, promotes transparent negotiations in the Brexit discussion	Green European Group	+
LINKE	written by left politician Sarah Wagenknecht, expresses harsh criticism on European Union and its economic policies, very polemic, very political	Die Zeit (Guest Article)	+
FDP	not accessible	-	-
AFD	discusses referendums as political instruments in general	Spiegel (Interview)	+
SPD	discusses racism and xenophobia in the context of the Brexit	Berliner Zeitung	+

Table 9: Brexit - Recommendation Content

9.3.2 Interpretation

The results of internal and external clustering validation measures showed different tendencies for the clustering of the Brexit articles. The highest silhouette coefficient for two clusters, can also be due to a bad distribution of samples across clusters. In the current case, the best silhouette coefficient is a result of one cluster with 23 articles and a second cluster with 1 article. However, we know that our samples are spread across the political spectrum, which makes it improbable that this result represents the actual perspectives. The explanation for this phenomenon is that individual outliers, lead to a good separation of the clusters and therefore the corresponding silhouette coefficient is still high.
Even though, the v-measure and the ARI favour our approach, the ARI showed rather poor results that do not allow to suggest the expected structure in the data. Overlapping clusters may be a reason for such bad ARI results. The recommendation results still reach the initial aim. We find that it is possible to recommend various articles on the same topic with little duplicate content, but high information gain.

9.4 Refugee Crisis and Right-Wing Populism

The experimental setting for the topics of the refugee crisis and right-wing populism is the same as for the Brexit articles. Table 8 shows the total amount of articles per topic: 30 for the refugee crisis and 28 for right-wing populism. The average text lengths are 534 words and 616 words. Both topics contain articles from seven different political parties. Therefore, seven is also the maximum amount of clusters we experiment with.

9.4.1 Results

The graphs in figure 9 show silhouette coefficients, v-measures and ARIs for the two topics.

Sihouette Coefficient:
The development of the silhouette coefficient differs from its development in the Brexit topic. For both topics, both clustering approaches bring the best results with the Manifesto labels. Maximum silhouette values are $s = 0.4361$ for K-Means clustering on Manifesto labels for the refugee crisis and seven clusters and $s = 0.6189$ for K-Means clustering on the Manifesto labels for the right-wing populism and five clusters. The K-Means clusterings of the BOW and the TF-IDF feature set perform better than the corresponding agglomerative clusterings.

V-Measure and Adjusted Rand Index:
While the highest value for the v-measure in the refugee crisis is a result of agglomerative clustering on TF-IDF features and seven clusters with $v = 0.4653$, right-wing populism reaches the highest value with K-Means clustering on the Manifesto labels and seven clusters with $v = 0.3858$. For both, this differs from the Brexit results.
Similar to the Brexit the ARI of the refugee crisis clustering has rather low values with a maximum of $ARI = 0.0996$ for K-Means clustering on Manifesto labels and four clusters. The right-wing populism only reaches an even lower ARI with a maximum value of $ARI = 0.0243$ for K-Means clustering on Manifesto labels. However, for this topic, the best performing approach is consistent with the v-measure and the silhouette results.

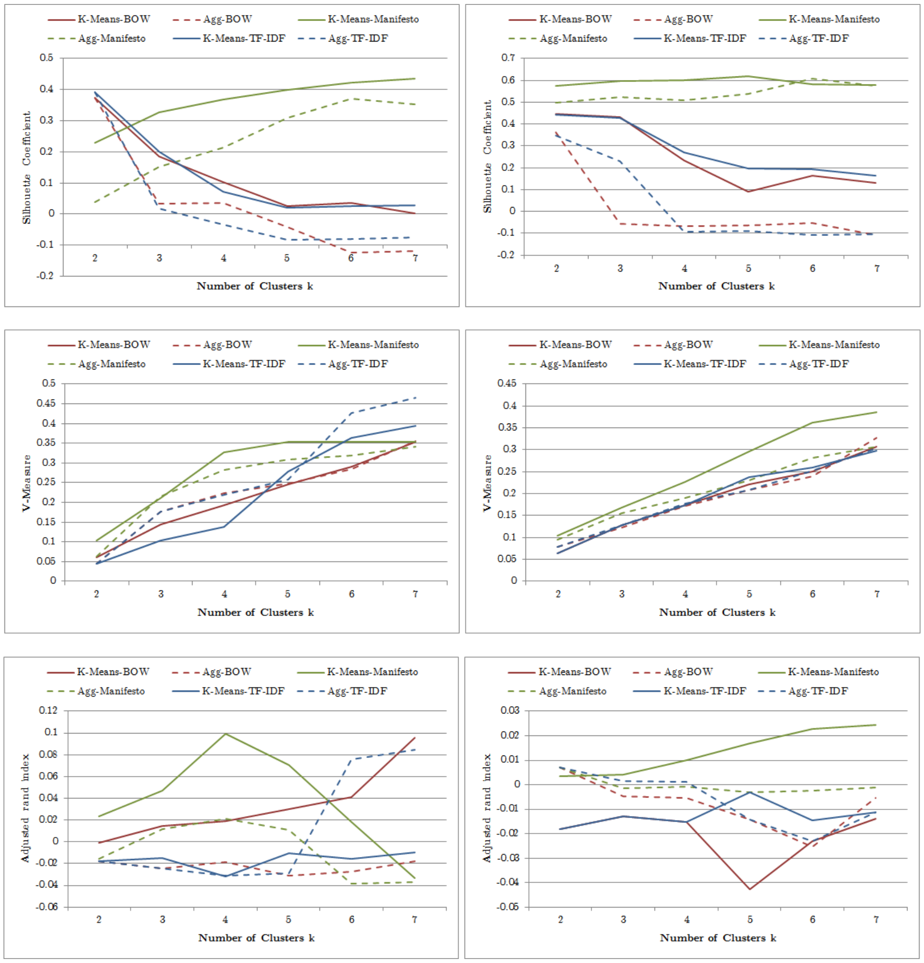


Figure 9: Refugee Crisis (left) and Right-Wing populism (right) - Silhouette, V-Measure, ARI

Recommendations:
Looking at the recommendations, we consider the set of recommended articles that comes from the approach with the highest ARI. For the refugee topic this is a K-Means-Manifesto combination and four clusters. We receive recommendations of articles that originate from the following political parties:

- **K-Means, Manifesto:** CSU, AFD, NPD, AFD

The recommendation set includes one double occurrence of the AFD. Furthermore, considering the traditional positioning of the political parties, the resulting parties have no representative of social democratic, left or green as they all belong to the center and right part of the political spectrum. Table 10 summarizes the content of the articles. Even though the articles have new aspects in

Origin	Content	Source	Inf.Gain
CSU	discusses consequences of German refugee policies, argument on contradicting statistics on the reasons of migration development between Merkel and Seehofer, further: welcome policy, blocking of the Balkan route and refugee deal with Turkey	Die Welt	+
AFD	informs about negotiations between the German Federal and State Governments on the costs of refugee integration	n-tv	+
NPD	reports on a fire in a refugee accommodation, claim that a refugee set the fire as a reaction on an argument about the provided food	Express	+
AFD	reports on the introduction of a law in Sweden that will result in the deportation of refugees	ParsToday	+

Table 10: Refugee Crisis - Recommendation Content

terms of content, the center/right position of the tweeting parties is reflected within their texts. The left side of a traditional political spectrum is missing. For the topic of the right-wing populism, we also analyze the articles of the K-Means-Manifesto clustering, as we found that it brought the best values for all three validation measures.

- **K-Means, Manifesto:** CSU, Grüne, DIE LINKE, CSU, SPD, FDP, SPD

Even though there are two double occurrences of political parties within the recommendations, they better cover the traditional political spectrum than in the previous topic. Again, we evaluate the content of the duplicate articles to see, whether they also provide duplicate content. Table 11 summarizes the content of the recommended articles. Again, the articles do not provide duplicate content, but supplementary information.

9.4.2 Interpretation

The results of these two topics were not consistent with the Brexit results. Still, the Manifesto labels performed best, regarding silhouette coefficients and ARI. Different results in the v-measure may also be a result of the limited amount of data. The documentation of the v-measure mentions that it might not bring good results with a small number of samples, with an example value of less than

Origin	**Content**	**Source**	**Inf.Gain**
CSU	reports that a CDU politician leaves the party for the right-wing party AFD	Badische Zeitung	+
Grüne	interview with Green politician Trittin on the Austrian president election, focus on the win of Green politician van der Bellen	Kölner Stadtanzeiger (Interview)	+
LINKE	cooperation between the right-wing parties AFD (German) and FPÖ (Austrian)	Handelsblatt	+
CSU	treats the role of Islam and Muslims in the German community, discusses, a questionable statement of a CSU politician, but depicts various perspectives on the topic, including left and green	Tagesschau	+
SPD	official SPD press release on the exclusion of an AFD member due to antisemitic statements, clearly expresses their fear that there may be more extreme thoughts in the party that have not yet reached the public	SPD Baden Würtemberg	+
FDP	FDP politician Schnarrenberger evaluates the results of the Austrian president election and the consequential risk for Germany, focus on AFD	Süddeutsche (Guest Article)	+
SPD	SPD politician Gabriel explaining the behavior of the AFD party: they abuse minorities as a justification for their arguments rather than looking for a discussion with the powerful parties. He mentions importance of open and content related discussions with the AFD.	Tagesspiegel	+

Table 11: Right-Wing Populism - Recommendation Content

a thousand samples. More particularly, it might not indicate random labeling with a value of zero [61]. We have far less samples than this in this experiment. The ARI accounts to this fact and also shows results close to zero, which would

imply random labeling and missing clustering structure. However, the silhouette coefficients suggest that there is a clustering structure in the data, which leads to the assumption that the labeling with political parties is not well suitable to our data.
As a difference to the Brexit, the definition of these two topics is very wide. The refugee crisis is in discussion since mid 2015, the right-wing populism even longer. Throughout this time period, there were various little events that arouse new arguments between political parties. Even though our articles come from a more limited time period of at most two months, all this former information may play a role in the news coverage as articles often refer to previous decisions, statements or events. The Brexit on the other hand was a very specific event and very recent as well. Therefore, the information on it is more concentrated. The refugee crisis and the right-wing populism do not refer to as specific events, they rather represent a development. Reports on these topics, might therefore split into more subtopics that again may carry different political perspectives. However, the refugee example shows that even if the articles represent subtopics of a high-level topic, they can still belong to the same political perspective, which is not intended. On the other hand, the right-wing populism had worse results, but covered a better political spectrum in its recommendations. Furthermore, the recommendations of the right-wing clustering show two striking features. The article of “Die LINKE” reports on the cooperation between two populist parties of the right-wing. The article only subliminally judges this event, but does not represent a clear left position. This highlights the weakness of the party labeling. Secondly, the second and the second last article (Grüne, SPD) report on the same event, the president election in Austria, but represent two different political perspectives very well. Also they both include direct speech of representative politicians.

9.5 Five Smaller Topics

We experiment with five more topics, for which we found between 10 and 20 articles. The topics are: Federal President Gauck (average 496 words), AFD politician Gauland (average 630 words), Safe countries of origin (average 671 words), the people’s party crisis and the German relation to Turkey (average 623 words). As the number of samples is low, we reduce the maximum number of clusters to five clusters and experiment with two to five clusters. The rest of the experiment setup remains as in the previous section. Table 8 shows the number of articles per topic (T4-T8) and the article distribution per party.

9.5.1 Results

In figures 10 and 11 we illustrate the results of the experiments.

Silhouette Coefficient:
Across all these five topics, we have the best silhouette values for the Manifesto feature set, with K-Means performing slightly better than agglomerative clus-

tering. For three out of five topics, the highest silhouette coefficients are above 0.5, for the other two they are above 0.3. While these results are consistent with the previous topics, they contradict with the external cluster validation measures.

V-Measure and Adjusted Rand Index:
For Federal president Gauck, v-measure and ARI show consistent results. The agglomerative clustering with the BOW and the TF-IDF feature set reach the highest values for both, v-measure and ARI and the maximum number of five clusters. With a v-measure of $v = 0.7394$ and an ARI of $ARI = 0.3514$ this topic brings the highest values overall.
For the next topic, AFD politician Gauland, the v-measure results and the ARI again show consistent results. Agglomerative clustering on TF-IDF vectors reaches the highest values for both, at five clusters. For the same amount of clusters agglomerative clustering on BOW-features and K-Means clustering on Manifesto features follow in rank. However, their ARI values are close to 0, while the agglomerative clustering on TF-IDF vectors reaches a value of $ARI = 0.2510$.
In the topic of safe countries of origin, K-Means on the BOW vectors and the TF-IDF vectors perform equally good and better than the other feature sets for v-measure and ARI and four clusters. The maximum v-measure is $v = 0.4956$ and the maximum ARI is $ARI = 0.2060$. The clusterings of the Manifesto labels follow them with an ARI of $ARI = 0.1339$.
For the party crisis topic, agglomerative clustering on TF-IDF features has the best performance with $v = 0.7424$ and $ARI = 0.3002$, followed by agglomerative clustering on BOW, with a much lower ARI of $ARI = 0.2111$. These values occur with the maximum number of five clusters.

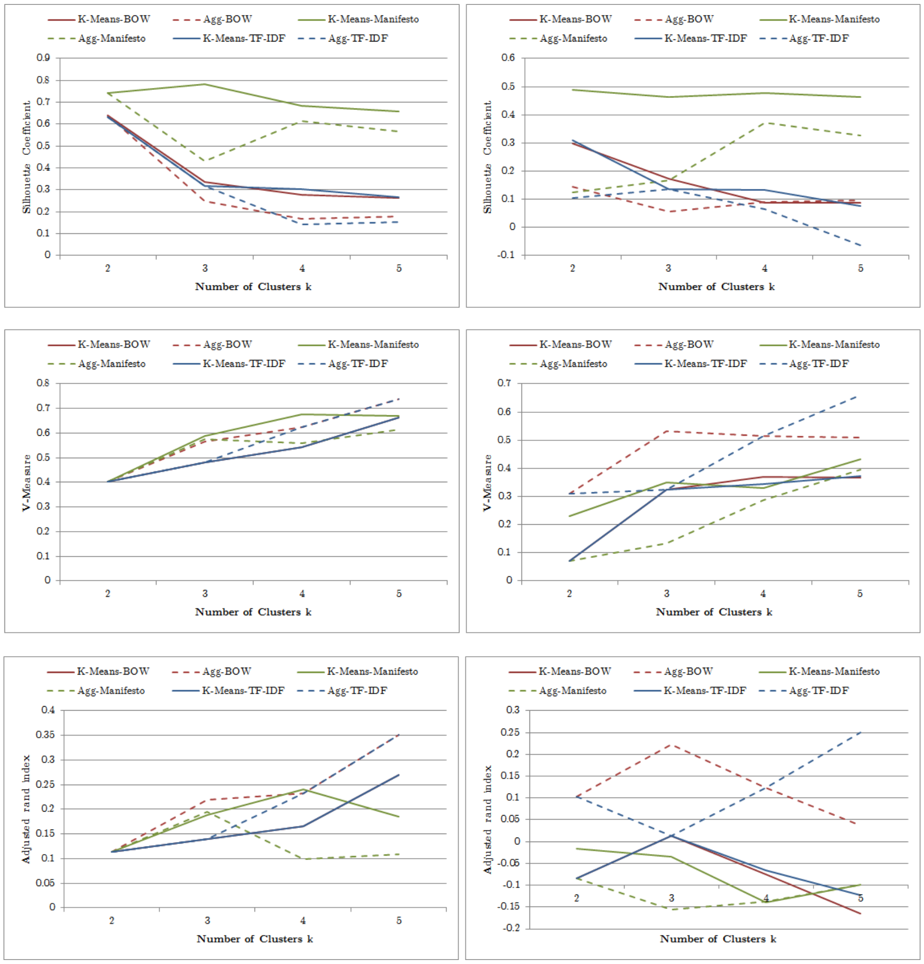


Figure 10: Federal President Gauck (left) and AFD Politician Gauland (right) - Silhouette, V-Measure, ARI

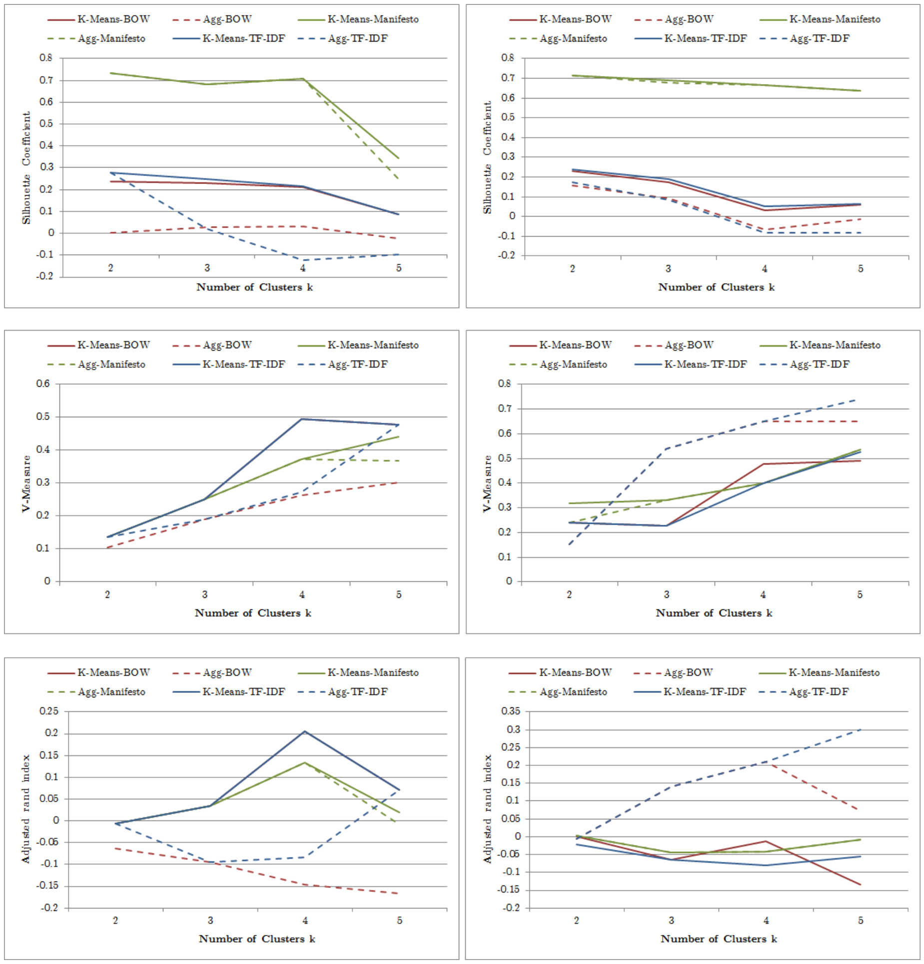


Figure 11: Safe Countries of Origin (left) and People's Party Crisis (right) - Silhouette, V-Measure, ARI

Regarding the topic of Turkey, for the v-measure, the TF-IDF features outperform the Manifesto labels and the BOW features with highest values for five clusters. This stands in contrast to the ARI, which has the highest value $ARI = 0.2006$ for agglomerative clustering on TF-IDF vectors with two clusters.

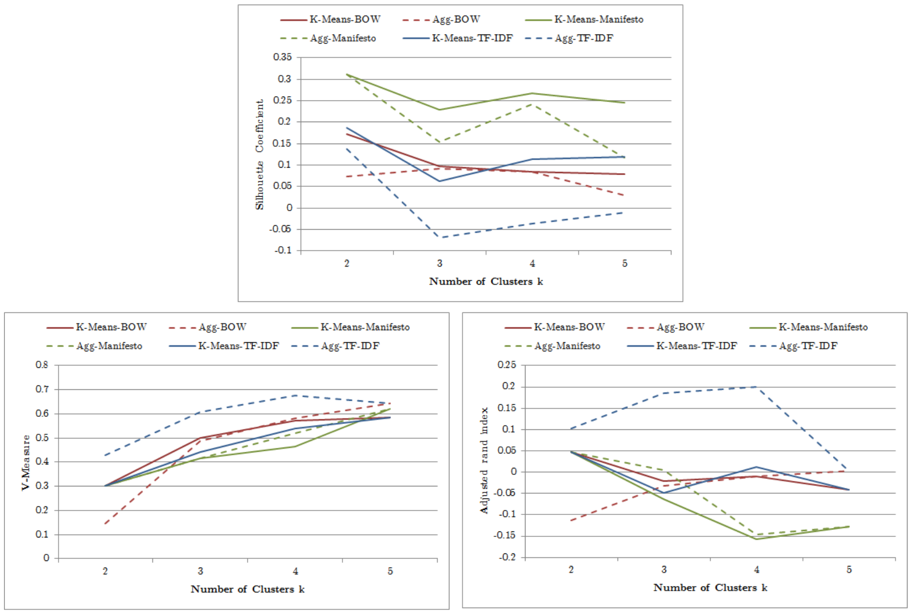


Figure 12: Turkey - Silhouette, V-Measure, ARI

Recommendations:
Regarding the results for Federal President Gauck, the topic stands out of the others as it is not very controversially discussed. Therefore, it was surprising to find rather clear structure within its articles, value-wise. Regarding the set of parties in the final recommendations, we found the following for agglomerative clustering on BOW and TF-IDF features:

- **TF-IDF, Aggl.:** CDU, FDP, AFD, DIE LINKE, DIE LINKE

Table 12 illustrates the recommendation content. The two articles from the left party clearly treat different topics, the others display the political perspective that corresponds to the originating party.
For the Gauland articles, the resulting recommendation contains articles that have been posted from the following parties:

- **TF-IDF, Aggl.:** SPD, Grüne, CDU, AFD, FDP

There are no double occurrences in this recommendation set.
For the safe states of origin the best ARI result, comes with the following set of reading recommendations:

- **TF-IDF, K-Means:** DIE LINKE, CDU, CDU, CDU

The result belongs to clustering with four clusters and within these, there are three re-occurrences of the CDU party. Table 13 displays the content of the recommendations. While these articles, all bring in some new aspects, there is

Origin	Content	Source	Inf.Gain
CDU	long article with different statements on the resignation for the next candidature period	NDR	+
FDP	Interview with FDP politician Lindner on the succession of president Gauck	NW (Interview)	+
AFD	succession plans from the AFD party	Merkur	+
DIE LINKE	critique on a fight between the CDU and the SPD party and the discussion from which party the following candidate will emerge	Morgenpost	+
DIE LINKE	expresses the preference for a candidate from the socialist/green/left side	Neues Deutschland	+

Table 12: Federal President Gauck - Recommendation Content

no article that actually represents and supports the opinion of the Green party, even though such articles were included in the set of articles. Therefore, we consider the distribution of the articles across the resulting clusters in Table 14:

We find that the distribution of articles per cluster is very unbalanced. While the first big cluster, contains all articles of the Green party as well as the Left, the CDU articles are spread across all four clusters. The Green and the Left party traditionally show agreement on certain political topics. This topic actually involves two main political perspectives: Being in favour of the declaration as safe countries or being against it. For reasons of comparability, we additionally evaluate the results for two and for eight clusters. Even with two clusters, we find one big cluster and another one with a single CDU article. In contrast, a clustering with 8 clusters breaks the big cluster down. However, this leaves seven clusters with only one article and seems to be to fine-grained
The clustering on the party crisis, with a higher ARI value ($ARI = 0.3002$), does include two party re-occurrences:

- **TF-IDF, Aggl.:** FDP, FDP, DIE LINKE, CDU, CDU

Nevertheless, it covers both sides of the traditional political spectrum. Table 15 illustrates its content. The articles bring in new content and different perspectives on the crisis. However, as this is a topic that affects the whole political landscape, it is unfavourable that neither social democrats, nor the Green party are represented by this recommendation set label-wise as well as content-wise
The recommendations for the topic of Turkey are as follows:

Origin	**Content**	**Source**	**Inf.Gain**
DIE LINKE	Report on Green politician Kretschmann's statement that he will agree on the declaration of Northern African countries as safe countries	Tagesschau	+
CDU	official statement of the CDU/CSU fraction that comments on the position that the Green party takes in the discussion. It is against this declaration.	CDU/CSU	+
CDU	how other political parties deal with the position of the Green party and whether they should compromise or take clear positions towards it	Hannoversche Allgemeine	+
CDU	very extensive article, expresses harsh criticism on the position of the Green party, claims that poverty-driven migration is no human right	Die Welt	+

Table 13: Safe Countries of Origin - Recommendation Content

- **TF-IDF, Aggl.:** DIE LINKE, FDP, FDP, CSU

We find one double occurrence of the FDP. Table 16 summarizes the content of the articles. This recommendation set covers multiple aspects of the relationship between Turkey and Germany. However, articles differ in their type and seriousness.

9.5.2 Interpretation

Again, we found that the Silhouette values are contradicting to the external validation measures. However this time, the ARI reaches higher values and does not imply random labeling. This is surprising as the number of articles is smaller now. We therefore expected worse results for this index. The differences in article type, especially for the last topic of Turkey show, that it might not be reasonable to compare too different typed articles. This challenge should be tackled in the article collection.

9.6 Label-Initialization

The K-Means clustering approach is sensitive to its initialization. In order to improve the results, we apply a better initialization approach for the clustering

#	2 Clusters	4 Clusters	8 Clusters
0	Grüne, Grüne, Grüne, CDU, CDU, CDU, Grüne, Grüne, DIE LINKE, DIE LINKE	Grüne, Grüne, Grüne, CDU, Grüne, Grüne, DIE LINKE, DIE LINKE	Grüne, Grüne, CDU, Grüne
1	CDU	CDU	DIE LINKE
2	-	CDU	CDU
3	-	CDU	CDU
4	-	-	DIE LINKE
5	-	-	CDU
6	-	-	Grüne
7	-	-	Grüne

Table 14: Party Distribution across Clusters for Safe Countries of Origin

Origin	Content	Source	Inf.Gain
FDP	poll on the agreement with the CSU in Bavaria, in which the CSU lost their absolute majority, in favour of the AFD party	Die Welt	+
FDP	official and detailed interview with FDP politician Lindner about the situation of his party, their goals and plans when they reenter parliament and the current political situation in Germany	Liberale.de (Interview)	+
DIE LINKE	comments on a very controversially discussed study ("Mitte-Studie")	Neues Deutschland	+
CDU	criticizes fights between the CDU and the CSU parties	Merkur	+
CDU	discusses the meaning of being conservative	Tagesspiegel	+

Table 15: Party Crisis - Recommendation Content

on Manifesto labels. We use the Manifesto label vectors for the political Manifestos/basic programmes of the political parties as initial centroids. Therefore, we partition the data into eight clusters corresponding to the eight political

Origin	Content	Source	Inf.Gain
DIE LINKE	criticism of Turkish reaction to the Armenia resolution	Tagesspiegel	+
FDP	comment on the refugee deal	Deutschland Radio Kultur	+
FDP	states 44 reasons to love Turkey	Bild Zeitung	+
CSU	postponement of the accession negotiations between Turkey and the European Union	Frankfurter Allgemeine Zeitung	+

Table 16: Turkey - Recommendation Content

parties.

9.6.1 Results

We compare the results of the improved initialization to K-Means clustering with the original initialization methods and eight clusters. Table 17 shows that silhouette coefficients are much higher for the k-means++ initialization. On average, they decrease by 49.6% with the application of the label initialization. However, for the V-Measure and ARI, the application of the new initialization results in an improvement of both values by 5.7%. The political spectrum covered by the two different recommendations, is as follows:

- **k-means++:** AFD, AFD, Grüne, AFD, CDU, FDP, SPD, Grüne
- **Label:** AFD, Grüne, Grüne, SPD, FDP, AFD, CDU, FDP

Both recommendation sets contain the same party spectrum, with differing articles. The spectrum pretty much represents the traditional political spectrum, leaving out the extreme edges.

9.6.2 Interpretation

The application of the label initialization did not improve the distribution of articles across the clusters, as initially intended. It also results in worse separation of the clusters, which is reflected in lower silhouette coefficients. It is possible that this is due to the fact that we partition the data into more clusters than actual political opinions. Also we do not have representative articles for every considered party. Therefore, it is very probable that clusters contain only a single article and that more than one cluster will represent a certain political direction, which explains worse cluster separation. On the other hand, this improves homogeneity values and thereby increases the v-measure. We consider this to be an overall positive effect on our method. However, it introduces new

Topic	**Initialization**	**Silhouette**	**V-Measure**	**ARI**
Brexit	k-means++	**0.6040**	0.4005	-0.0371
	Label	0.1285	**0.4227**	**-0.0255**
Refugees	k-means++	**0.4554**	0.3775	-0.0391
	Label	0.3434	**0.4546**	**0.0370**
Right-Wing Populism	k-means++	**0.5970**	0.3847	0.0069
	Label	0.4629	**0.4263**	**0.353**
Gauck	k-means++	**0.4514**	0.7189	0.0228
	Label	0.2978	0.7189	0.0228
Gauland	k-means++	**0.3735**	0.5849	-0.0450
	Label	0.1706	**0.6733**	**0.1222**
Safe Countries of Origin	k-means++	**0.2093**	0.5132	-0.0806
	Label	0.1293	**0.5504**	**-0.0185**
Party Crisis	k-means++	**0.3261**	0.7807	0.0859
	Label	0.0347	**0.7895**	**0.1081**
Turkey	k-means++	**0.1811**	**0.8069**	**-0.0630**
	Label	0.0445	0.7905	-0.0825

Table 17: K-Means++ vs. Label Initialization for k=8

challenges, as we will have to restrict the number of articles for the recommendation. A way to do this would be to choose the clusters with the best individual silhouette coefficients.

9.7 Cross-Topic Clustering

As a side experiment we analyze whether the language of parties differs across various topics. More specifically, we aim to answer the question whether it is possible to identify a party by the language of the news articles they tweet, independent of the actual topic of that news article. This may be hard to achieve as there is usually an additional communication layer between the politicians' speech and the news articles. Politicians usually do not write or speak directly in their tweeted articles, but express their opinion by the speech of the journalists. For this scenario, we extract all article texts across all topics (168 articles at that point of time). As before, we evaluate K-Means and Hierarchical clustering with the three feature sets, accordingly. As we represent eight parties through the articles, we experiment with seven, eight and nine clusters as a small range around these eight actual categories. As a comparative experiment, we cluster the articles again, but compare the result to their topic rather than their political label. As the TF-IDF measure is a common instrument to detect topics of texts [62], for example to find the most relevant documents to a search query in a search engine, we expect the TF-IDF labels to show the best performance for this experiment. We have 22 topics overall. A lot of them are related to each other. As an example, Right-Wing populism, AFD politician Gauland and the people's crisis do overlap in certain aspects. Likewise, Turkey and the refugee crisis have common aspects of discussion as well. To respect these overlaps we experiment with 13 to 22 topics in this experiment The articles have an average length of 567 words.

9.7.1 Results

As before we evaluate the silhouette coefficient, the v-measure and the ARI for this experiment. Figure 13 shows the results. Similar to the previous experiments, clustering the Manifesto labels results in the highest silhouette coefficients. K-Means reaches the top score for both reference labels, the political parties and the topics with values > 0.5 with agglomerative clustering on the TF-IDF vectors. For the v-measure, the political parties have a much lower maximum value ($v = 0.1676$) than the topic labels ($v = 0.7077$). The same is true for the ARI with maximum values of $ARI = 0.0241$ for the political party reference and $ARI = 0.3778$ for the topic labels.

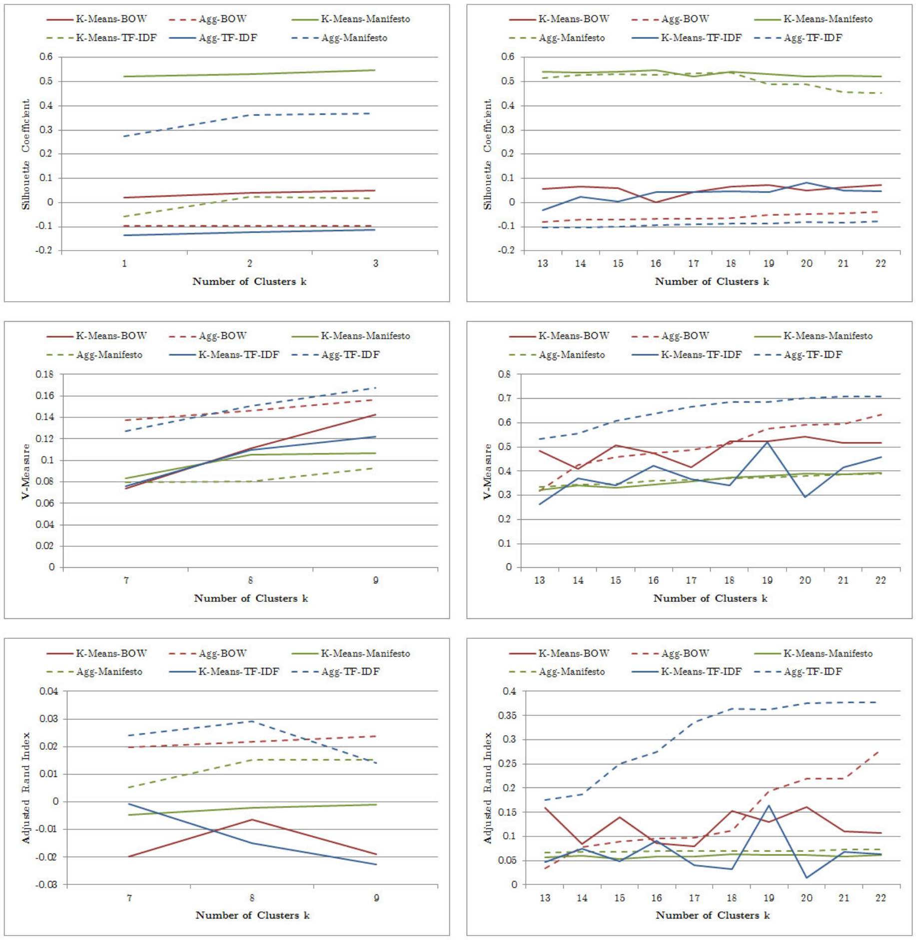


Figure 13: Cross-Topic Clustering- Silhouette, V-Measure, ARI
Party Labels (left) and Topics (right)

9.7.2 Interpretation

The results confirm that TF-IDF vectors reflect the topic of a text as we have much higher values if we consider the topics as reference labels. Further, the results imply that political directions cannot be extracted from cross-topic texts with the Manifesto labels. If we consider the political parties as reference labels, the values for the ARI are very close to 0. This proves that there is less structure in the data than if we consider the articles per topic as in the previous experiments. For the TF-IDF measure as well as the BOW-featuers this is not surprising, as they are directly related to the actual words of a text. It is obvious that the vocabulary across different topics shows more variation than the vocabulary of texts that originate from the same topic. But, as the Manifesto labels represent ideologies and not topics, we expected them to better represent cross-topic texts. Still, the different topics can represent different ideologies each, if we consider the labels, such as capitalism, military, the European financial crisis will probably have higher values for positive and negative capitalism, while articles on refugees or the related Syrian war, will focus on military aspects, rather than capitalistic ones. Therefore, we can conclude that the Manifesto labels as characterization criterion correlate with the topics of a text, but do not solely represent a political direction on a cross-topic basis.

9.8 General Obeservations across all Experiments

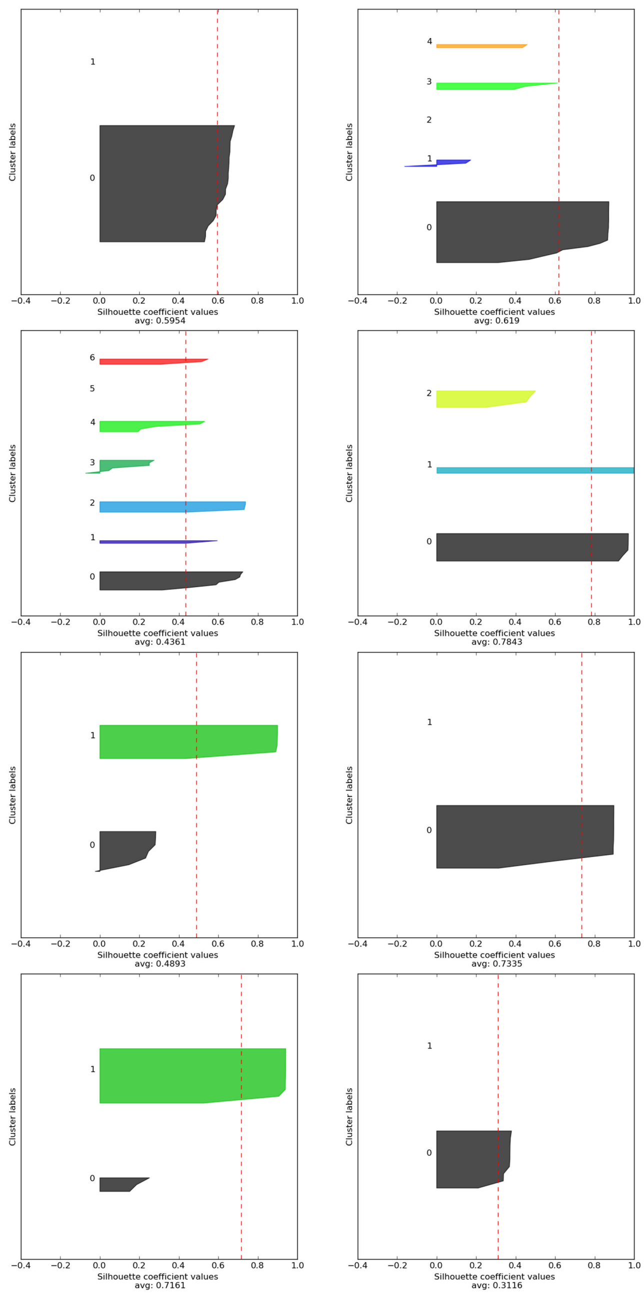
As the results across the different topics differ and are not totally consistent, this section gives an overview on how the different approaches behaved across all experiments and leads to the final conclusions of this work. Therefore, we consider average values of the approaches as well as general trends.
Overall, we evaluated our approach on eight different topics. As previous individual analyses show, the external and internal cluster validation measures were not consistent to each other throughout the majority of the experiments.
The results of internal cluster validation with the silhouette coefficient were mostly consistent across all eight topics. Table 18 shows the best combinations of feature sets and clustering approaches in respect to average silhouette coefficients. Except for the Brexit, where the BOW reached the highest silhouette coefficient for both K-Means and agglomerative clustering, all other topics have the best results for K-Means clustering on the Manifesto labels, followed by agglomerative clustering on the Manifesto labels. This corresponds to the expected results, as the Manifesto labels should represent political perspectives better than the other two feature sets. Within the other two feature sets, K-Means clusterings outperformed agglomerative clusterings of both sets. That K-Means in general reaches higher silhouette coefficients than agglomerative clustering is also due to the calculation of the measure. It is based on Euclidean distance and therefore favors K-Means, as its implementation also bases upon the same distance metric. Therefore, we do not consider the silhouette coefficient as an appropriate instrument for the comparison of the two different cluster approaches. Nevertheless, it is valuable for the comparison of the different feature sets, as well as the amount of clusters for each clustering, bearing in mind the risk of manipulated values through outliers. The optimal amount varies between two clusters for Brexit, Gauland, safe countries of origin, party crisis and seven clusters for the refugee topic. The optimal number of clusters does generally not correspond to the number of parties.

Regarding the values of the maxima, they vary between $s = 0.3103$ and $s = 0.7843$, which corresponds to a low structure in the data up to a strong and clear clustering structure in the data. However, it is important to not only consider this average silhouette value individually but also evaluate the visual representation of the resulting silhouettes for each cluster. Clusterings with individual outliers can also result in high average silhouette coefficients, as we saw in the topic of safe states of origin. In this case, there is one huge cluster and another cluster that contains the outlier. This results in high silhouette

Topic	Feature Set	Clustering Approach	Value
Brexit	BOW	Agglomerative & K-Means	0.5954
Right-Wing Populism	Manifesto	K-Means	0.6190
Refugees	Manifesto	K-Means	0.4361
Gauck	BOW & Manifesto	K-Means	0.7843
Gauland	Manifesto	K-Means	0.4769
Secure States	Manifesto	Agglomerative & K-Means	0.7335
Party Crisis	Manifesto	K-Means	0.7095
Turkey	Manifesto	Agglomerative & K-Means	0.3116

Table 18: Best Performing Approaches for the Silhouette Coefficient

coefficients as the huge cluster is very well separated from the outlier [54]. We display the silhouettes per cluster in figure 14 for the approach that reached the maximum average silhouette coefficient in each topic. Particularly for the topics, which reach the highest average silhouette values with only two clusters, we observe that there is one very big cluster and a second cluster, which contains only one article (their silhouette is not visible at all). This is the case for the Brexit silhouettes, as well as the silhouettes for safe countries of origin and Turkey. In contrast, the refugee crisis has rather equally sized clusters. This observation implies that even though the maximum average silhouette value may be reached at two clusters only, a higher number of clusters may actually represent a better clustering, with more equal sized clusters, even with a lower silhouette coefficient. Figure 15 confirms this suggestion. There are two silhouette visualizations for the topic of the Brexit. The visualization of the two clusters that result from agglomerative clustering on the BOW vectors and the six silhouettes of the K-Means clustering on the Manifesto labels. Even though we have the highest value for the BOW clustering, the result are two very unequal clusters. In the Manifesto clustering the samples are better distributed across the clusters. This shows that a two-dimensional view on the articles would not be appropriate, which again encourages our decision not to rely on the initial left-right classification of the Manifesto classifier. Furthermore, it explains inconsistencies between internal and external cluster validation measures. It also implies that there is no clear distinction of a certain number of political perspectives. Overlapping clusters and perspectives, respectively, may result in lower values for external validation measures as well. The fact that we found the same news articles on Twitter profiles of different political parties strengthens this suggestion.



Brexit (left), Right-Wing Populism (right)

Refugee Crisis(left), Federal President Gauck (right)

AFD Politician Gauland(left), Safe Countries of Origin (right)

Party Crisis(left), Turkey (right)

Figure 14: Silhouette Visualization for max Silhouette Values

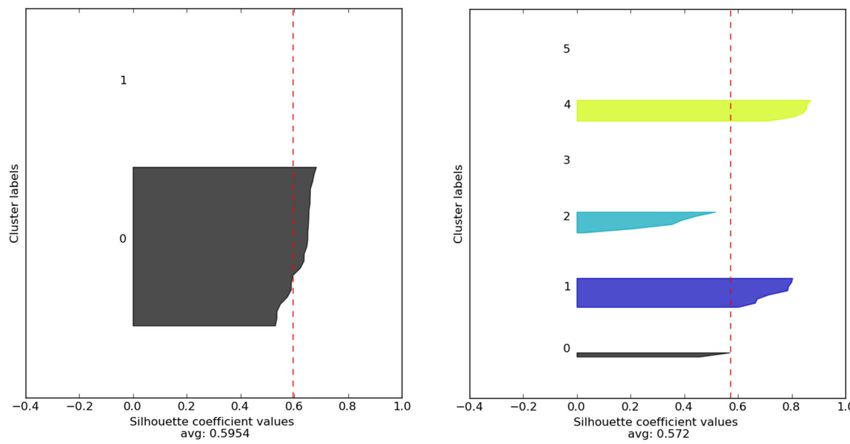


Figure 15: Silhouette Visualization for the Brexit
Agglomerative on BOW (left), K-Means on Manifesto (right)

The external validation measures v-measure and ARI rely on the reference categorization of the tweeting political parties. Table 19 shows the best performing combinations of features and clustering approaches for all eight topics, regarding the v-measure. In this case, clustering on TF-IDF features outperforms Manifesto clusterings in six out of eight topics. Furthermore, agglomerative clustering results in higher values for the v-measure. The values' range is between $v = 0.3560$ and $v = 0.7424$, where $v = 1.0$ would represent a perfect match between the original labels and the clustering results. However, we do not always have as many clusters as we have political parties in our reference. Therefore, it would not always be possible to reach a value of 1. But the v-measure considers homogeneity and completeness. High completeness can be reached even with lower cluster numbers, which is why we still consider the v-measure to be an appropriate measure for our validation.
Table 20 displays the best feature and approach combinations regarding the ARI. They are mostly consistent with the previous results of the v-measure. In three out of eight topics the clustering on the Manifesto labels outperformed the clustering of TF-IDF feature set. This corresponds to 37.5%. The other five topics bring the best result for the clustering on TF-IDF vectors (62.5%). Agglomerative clustering performs better in 62.5%. In general, the values of the ARI are not very high, varying between $ARI = 0.0178$ and $ARI = 0.3514$. Values close to zero indicate random labeling. In this context, we observe that the five "small" topics that have datasets with less than twenty samples have more than ten times higher values than the three bigger data sets. [53] mention that the ratio between the number of samples and the number of clusters has an impact on the outcome of the Rand index. Likewise, this might be the case for the adjusted version of this measure.

Apart from quantitative results, the qualitative evaluation of the recommendations showed that we experience an information gain for all topics. We found

Topic	Feature Set	Clustering Approach	Value
Brexit	Manifesto	Agglomerative	0.3560
Right-Wing Populism	Manifesto	K-Means	0.3858
Refugees	TF-IDF	Agglomerative	0.4653
Gauck	BOW & TF-IDF	Agglomerative	0.7394
Gauland	TF-IDF	Agglomerative	0.6612
Secure States	BOW & TF-IDF	K-Means	0.4956
Party Crisis	TF-IDF	Agglomerative	0.7424
Turkey	TF-IDF	Agglomerative	0.6778

Table 19: Best Performing Approaches for the V-Measure

Topic	Feature Set	Clustering Approach	Value
Brexit	Manifesto	Agglomerative	0.0178
Right-Wing Populism	Manifesto	K-Means	0.0243
Refugees	Manifesto	K-Means	0.0996
Gauck	BOW & TF-IDF	Agglomerative	0.3514
Gauland	TF-IDF	Agglomerative	0.2510
Secure States	BOW & TF-IDF	K-Means	0.2060
Party Crisis	TF-IDF	Agglomerative	0.3002
Turkey	TF-IDF	Agglomerative	0.2006

Table 20: Best Performing Approaches for the Adjusted Rand Index

that rather than duplicate content, missing political perspectives pose a bigger challenge. Furthermore, we figure that a full coverage of the political spectrum in the recommendation set is more relevant than having individual representatives of each political party. For the Brexit, right-wing populism, Gauck, Gauland and Turkey the traditional political spectrum was well covered. For the refugee crisis, the party crisis and the safe countries of origin we found a part of the spectrum to be missing. Furthermore, different types of articles may result in difficult comparability as well. It is hard to compare a long, extensive interview with a short provocative article. Even though they may not duplicate content the information gain differs drastically. In general, we find that interviews with politicians are very helpful for a representation of a political opinion.

10 Conclusion and Future Work

We present a novel approach for a news recommendation system that illustrates different political perspectives on controversially discussed news topics. For this purpose, we suggest a content-based recommendation approach that uses automated text analysis to extract features that define political perspectives and then recommends a set of articles on a political topic that incorporates these perspectives. Based on former literature evaluation, this approach comes with many advantages over others, such as collaborative filters or personalized content-based recommendations. It accounts to the fact that most news websites do not require user profiles for news consumption as well as the very dynamic news environment, which in case of non-automated approaches would lead to significant latencies regarding recommendations. Moreover, it overcomes the obstacle of required expert knowledge for the item characterization.

The text analysis is the core of the model. We apply a clustering approach on the extracted text features in order to partition the articles into their belonging perspectives. We further suggest to use the domain-specific Manifesto classifier that extracts political ideologies from a given text. Within the three feature sets that we compare, we find that the Manifesto labels and TF-IDF vectors achieve better results than BOW vectors. In regard to our second research question that is directed to a suitable clustering approach for the Manifesto label space, we find that in terms of the ARI validation measure, K-Means outperformed agglomerative clustering with the cosine distance metric and the average linkage criterion for the Manifesto label set. The same is true, considering the average silhouette coefficients as a validation criterion. Regarding the clustering on TF-IDF vectors, we reached better results with agglomerative clustering. This is in line with former research approaches.

Even though the common TF-IDF vectors show superior results for some topics, they come with an important disadvantage in regard to future applications. While it is possible to directly calculate Manifesto labels for each new item straight away, the inverse document frequency of the TF-IDF calculation requires knowledge on all documents in a data set. For a real-life application this means that all TF-IDF vectors would have to be recalculated whenever a new item of an existing topic arrives. The same applies to the BOW features. As TF-IDF vectors are common for topic extraction these results lead to the assumption that topics and perspectives correlate. We further infer that due to the restriction to the lexical subspace for a topic, TF-IDF is also applicable to identify more subtle differences between texts.

The manual recommendation evaluation showed little duplicate content in the recommendation sets. Most articles, as intended, bring an information gain and deeper insight into a topic. However, we consider missing political perspectives to be a yet unsolved challenge in our approach. Concerning the debate on the refugee crisis and the safe states of origin, we do not achieve a comprehensive coverage of the political spectrum. While the refugee topic brought recommendations that were settled in the center and right part of the political spectrum, the safe states of origin showed representations of the Left party and the CDU. In this context, we realize that taking into account extreme political positions is not necessarily adding value as they risk to drift to non-objective argumentations.

We also observe that the type of article has an important impact, which we did not consider in our model. To represent rather clear political perspectives, interviews with the representative politicians of different parties or guest articles are a common as well as very informative type of article. Within these texts, politicians often describe and justify their view on certain topics. Furthermore, they include, at least in parts, direct speech of politicians, which may reduce distortion biases that originate from journalists. Their texts are usually quite long, which is preferable for feature extraction and can hardly be compared to very short, factual articles. In consequence, we suggest to follow up on this work, using more homogeneous texts in terms of type and length.

Our work comes not without limitations. Twitter has become an important communication channel in politics. Therefore, articles originating from Twitter accounts have the potential to express opinions very well. However, it is hard to find a large number of articles per topic that represent various political perspectives. Our clustering validation methods, particularly the v-measure and the related completeness and homogeneity measure would benefit from bigger sample sets by producing more reliable results. To overcome this, we focused on the adjusted Rand index as validation criterion. Moreover, an evaluation of more topics would proof the general validity of our findings. Ideally, a new data set for further development of this approach would include more than 50 articles per topic that have been politically classified by experts. To our knowledge no such dataset exists. This leads to the second challenge in this task: the definition of different political perspectives. We found that there are various ways to differentiate political perspectives, starting from a two-dimensional left/right representation up to a perspective for each political party. However, they are not always clearly separable. As for now, we suggest to aim for a coverage of the traditional political spectrum. Opinions are dynamic and parties develop over time. In this context, fuzzy clustering might also be a method that is able to capture these characteristics.

In clustering, validation measures represent a way to give a hint on the quality of the clustering, but are not sufficient for a final judgement [22]. We experience this as our applied validation measures do not always show consistent results from one to another. While they are helpful and informative in such an experimental stage of the work, a way to get around these validation weaknesses, is to pursue a user survey to empirically evaluate the recommendation quality. We leave this for future work.

In the scope of this work, we particularly focused on the applicability of the Manifesto labels as an instrument of text characterization for clustering as well as the implementation of a suitable clustering approach. On the way towards a fully developed recommendation engine for multi-perspective understanding of political topics further aspects need to be addressed: First of all, the collection of news items: It is very important to ensure source diversity, regarding the considered items. Even for future use Twitter represents an interesting news source. Political profiles could be crawled regularly. Thereby, it is necessary to ensure to only consider actual news items as there is a lot of other unrelated content, such as videos or updates on politicians' activities. The detection of relevant and controversially discussed political topics is another aspect that we covered only manually. Hash tags already represent a means to group topic related tweets together. Furthermore, Twitter provides information about currently hot topics in news. Nevertheless, not all these topics are subject to controversial discussions. Therefore, a topic filter that takes into consideration the political impact and the scope of a topic should be implemented.

Until now, we only considered German news and the German political landscape in our approach. For the future, it would certainly be interesting to extend this work to further countries. As the Manifesto corpus comprises political manifestos from all over the world, the classifier could easily be adapted to other languages. While Germany, as well as most European countries, enjoys freedom of press, this is not the case everywhere. In countries with censorship, people suffer from restrictions that are much worse than the filter bubble or media bias. In this context, a cross-national approach that compares and opposes political press across different countries would be of particular value.

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. Personalization technologies: a process-oriented perspective. *Communications of the ACM*, 48(10): 83–90, 2005.

[2] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.

[3] Yannis Bakos. Reducing buyer search costs: Implications for electronic marketplaces. *Management science*, 43(12):1676–1692, 1997.

[4] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[5] James Bennett and Stan Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35, 2007.

[6] Felix Biessmann, Pola Lehmann, Daniel Kirsch, and Sebastian Schelter. Predicting political party affiliation from text.

[7] Daniel Billsus and Michael J. Pazzani. A personal news agent that talks, learns and explains. *Proceedings of the third annual conference on Autonomous Agents*, pages 268–275, 1999. doi: 10.1145/301136.301208. URL `http://portal.acm.org/citation.cfm?id=301208{&}dl=GUIDE,`.

[8] Daniel Billsus, Michael J. Pazzani, and James Chen. A learning agent for wireless news access. *Proceedings of the 5th international conference on Intelligent user interfaces - IUI '00*, pages 33–36, 2000. doi: 10.1145/325737.325768. URL `http://portal.acm.org/citation.cfm?id=325737.325768$\delimiter"026E30F$nhttp://portal.acm.org/citation.cfm?doid=325737.325768`.

[9] Toine Bogers and Marijn Koolen. Report on recsys 2015 workshop on new trends in content-based recommender systems. In *ACM SIGIR Forum*, volume 49, pages 141–146. ACM, 2016.

[10] Engin Bozdag, Qi Gao, Geert-Jan Houben, and Martijn Warnier. Does offline political segregation affect the filter bubble? an empirical analysis of information diversity for dutch and turkish twitter users. *Computers in Human Behavior*, 41:405 – 415, 2014. ISSN 0747-5632. doi: http://dx.doi.org/10.1016/j.chb.2014.05.028. URL `http://www.sciencedirect.com/science/article/pii/S0747563214003069`.

[11] Erik Brynjolfsson, Yu Jeffrey Hu, and Michael D Smith. From niches to riches: Anatomy of the long tail. *Sloan Management Review*, 47(4):67–71, 2006.

[12] Michel Capelle, Flavius Frasincar, and Marnix Moerland. Semantics-Based News Recommendation. *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, pages 27–36, 2012. doi: 10.1145/2254129.2254163. URL `http://dl.acm.org/citation.cfm?id=2254163`.

[13] Michel Capelle, Marnix Moerland, Frederik Hogenboom, Flavius Frasincar, and Damir Vandic. Bing-sf-idf+: a hybrid semantics-driven news recommender. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pages 732–739. ACM, 2015.

[14] Yoonjung Choi, Yuchul Jung, and Sung-Hyon Myaeng. *Intelligence and Security Informatics: Pacific Asia Workshop, PAISI 2010, Hyderabad, India, June 21, 2010. Proceedings*, chapter Identifying Controversial Issues and Their Sub-topics in News Articles, pages 140–153. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-13601-6. doi: 10.1007/978-3-642-13601-6_16. URL `http://dx.doi.org/10.1007/978-3-642-13601-6_16`.

[15] Michael Conover, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. *ICWSM*, 133:89–96, 2011.

[16] Alexander Dallmann, Florian Lemmerich, Daniel Zoller, and Andreas Hotho. Media bias in german online newspapers. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, pages 133–137, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3395-5. doi: 10.1145/2700171.2791057. URL `http://doi.acm.org/10.1145/2700171.2791057`.

[17] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. Linked open data to support content-based recommender systems. In *Proceedings of the 8th International Conference on Semantic Systems*, pages 1–8. ACM, 2012.

[18] Robert M Entman. Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173, 2007.

[19] Bundeszentrale für politische Bildung. Wahl-O-Mat. `http://www.bpb.de/politik/wahlen/wahl-o-mat/166945/wie-funktioniert-der-wahl-o-mat/`, 2013. [Online; accessed 31-May-2016].

[20] R Kelly Garrett and Paul Resnick. Resisting political fragmentation on the internet. *Daedalus*, 140(4):108–120, 2011.

[21] Jörg Gottschlich, Irina Heimbach, Oliver Hinz, et al. The value of users' facebook profile data-generating product recommendations for online social shopping sites. In *ECIS*, page 117, 2013.

[22] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. On clustering validation techniques. *Journal of intelligent information systems*, 17(2-3): 107–145, 2001.

[23] Oliver Hinz and Jochen Eckert. The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering*, 2(2):67–77, 2010.

[24] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985. ISSN 1432-1343. doi: 10.1007/BF01908075. URL http://dx.doi.org/10.1007/BF01908075.

[25] Jon Espen Ingvaldsen, Özlem Özgöbek, and Jon Atle Gulla. Context-aware user-driven news recommendation. In *Proceedings of the 3rd International Workshop on News Recommendation and Analytics (INRA 2015) co-located with 9th ACM Conference on Recommender Systems (RecSys 2015), Vienna, Austria, September 20, 2015.*, pages 33–36, 2015. URL http://ceur-ws.org/Vol-1542/paper5.pdf.

[26] Stefano Kaplan, Ethan and DellaVigna. THE FOX NEWS EFFECT : MEDIA BIAS AND VOTING by Stefano DellaVigna and Ethan Kaplan Stockholm University. *Quarterly Journal of Economics*, 2007.

[27] Benjamin Kille, Andreas Lommatzsch, Roberto Turrin, András Serény, Martha Larson, Torben Brodt, Jonas Seiler, and Frank Hopfgartner. Overview of clef newsreel 2015: News recommendation evaluation lab. 2015.

[28] Yehuda Koren and Robert Bell. Advances in collaborative filtering. In *Recommender systems handbook*, pages 145–186. Springer, 2011.

[29] Steven Kull, Clay Ramsay, and Evan Lewis. Misperceptions, the media, and the iraq war. *Political Science Quarterly*, 118(4):569–598, 2003.

[30] Victor Lavrenko, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. Language Models for Financial News Recommendation. *Proceedings of the 9th International Conference on Information and Knowledge Management*, pages 389–396, 2000. doi: 10.1145/354756.354845. URL http://citeseer.ist.psu.edu/373966.html$\delimiter"026E30F$nhttp://portal.acm.org/citation.cfm?id=354845.

[31] Lei Li, Dingding Wang, Tao Li, Daniel Knox, and Balaji Padmanabhan. SCENE : A Scalable Two-Stage Personalized News Recommendation System. pages 125–134, 2011.

[32] Shiu li Huang. Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods. *Electronic Commerce Research and Applications*, 10(4):398 – 407, 2011. ISSN 1567-4223. doi: http://dx.doi.org/10.1016/j.elerap.2010.11.003. URL http://www.sciencedirect.com/science/article/pii/S156742231000089X.

[33] Q Vera Liao and Wai-Tat Fu. Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2359–2368. ACM, 2013.

[34] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 31–40. ACM, 2010.

[35] Martin Löffler. Der verfassungsauftrag der publizistik. *Publizistik (1960)*, (6):517–522, 1960.

[36] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*, pages 73–105. Springer, 2011.

[37] Ismini Lourentzou. Hopspots of News Articles : Joint Mining of News Text and Social Media to Discover Controversial Points in News. 2012.

[38] Paul Lucardie. *Zur Typologie der politischen Parteien*, pages 61–76. Springer Fachmedien Wiesbaden, Wiesbaden, 2013. ISBN 978-3-658-00963-2. doi: 10.1007/978-3-658-00963-2_2. URL `http://dx.doi.org/10.1007/978-3-658-00963-2_2`.

[39] Christian Matt, Alexander Benlian, Thomas Hess, and Christian Weiß. Escaping from the filter bubble? the effects of novelty and serendipity on users' evaluations of online recommendations. In *Proceedings of the International Conference on Information Systems - Building a Better World through Information Systems, ICIS 2014, Auckland, New Zealand, December 14-17, 2014*, 2014. URL `http://aisel.aisnet.org/icis2014/proceedings/HumanBehavior/33`.

[40] Brian McNair. *Journalism and democracy*. London: Routledge, 2000.

[41] Denis McQuail. *McQuail's mass communication theory*. Sage publications, 2010.

[42] Nicolas Merz, Sven Regel, and Jirka Lewandowski. The manifesto corpus: A new resource for research on political parties and quantitative text analysis. *Research & Politics*, 3(2):2053168016643346, 2016.

[43] D. Michie, D. J. Spiegelhalter, and C.C. Taylor. Machine learning, neural and statistical classification, 1994.

[44] D. Mladenic. Text-learning and related intelligent agents: a survey. *IEEE Intelligent Systems and their Applications*, 14(4):44–54, Jul 1999. ISSN 1094-7167. doi: 10.1109/5254.784084.

[45] Frank Nullmeier. Links–rechts. *Stephan Lessenich/Frank Nullmeier (Hg.): Deutschland. Eine gespaltene Gesellschaft. BpB: Bonn*, pages 313–335, 2006.

[46] Newspaper Association of America. Newspaper Digital Audience Grew Twice as Fast as the Internet in Past 12 Months. `http://www.naa.org/~/media/NAACorp/Public%20Files/TrendsAndNumbers/Newspaper-Websites/Final_Aug-2015_DigitalAudience.pdf`, 2015. [Online; accessed 29-July-2016].

[47] Kevin McCarthy & Barry Smyth Owen Phelan, Owen Phelan, Kevin McCarthy, and Barry Smyth. Using Twitter to Recommend Real-Time Topical News. *Acm*, pages 385–388, 2009. doi: 10.1145/1639714.1639794. URL `http://portal.acm.org/citation.cfm?id=1639794`.

[48] Matteo Palmonari, Giorgio Uboldi, Marco Cremaschi, Daniele Ciminieri, and Federico Bianchi. Dacena: Serendipitous news reading with data contexts. In *The Semantic Web: ESWC 2015 Satellite Events*, pages 133–137. Springer, 2015.

[49] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[50] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. Newscube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–452. ACM, 2009.

[51] Anand Rajaraman, Jeffrey D Ullman, Jeffrey David Ullman, and Jeffrey David Ullman. *Mining of massive datasets*, volume 1. Cambridge University Press Cambridge, 2012.

[52] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Introduction to recommender systems handbook*. Springer, 2011.

[53] Andrew Rosenberg and Julia Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. In *EMNLP-CoNLL*, volume 7, pages 410–420, 2007.

[54] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987. ISSN 0377-0427. doi: http://dx.doi.org/10.1016/0377-0427(87)90125-7. URL `http://www.sciencedirect.com/science/article/pii/0377042787901257`.

[55] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5): 513–523, 1988.

[56] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[57] J Ben Schafer, Joseph A Konstan, and John Riedl. E-commerce recommendation applications. In *Applications of Data Mining to Electronic Commerce*, pages 115–153. Springer, 2001.

[58] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

[59] scikit learn. Linkage Criterion for Agglomerative Clustering. `http://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html`, 2013. [Online; accessed 30-July-2016].

[60] scikit learn. TfidfTransformer. `http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfTransformer.html#id1`, 2013. [Online; accessed 23-July-2016].

[61] scikit learn. V-Measure. `http://scikit-learn.org/stable/modules/clustering.html#homogeneity-completeness`, 2013. [Online; accessed 27-July-2016].

[62] Kristie Seymore and Roni Rosenfeld. Using story topics for language model adaptation. 1997.

[63] Michael Steinbach, George Karypis, Vipin Kumar, et al. A comparison of document clustering techniques. In *KDD workshop on text mining*, volume 400, pages 525–526. Boston, 2000.

[64] George J Stigler. The economics of information. *The journal of political economy*, pages 213–225, 1961.

[65] Sergios Theodoridis and Konstantinos Koutroumbas. Chapter 16 - cluster validity. In Sergios Theodoridis, , and Konstantinos Koutroumbas, editors, *Pattern Recognition (Fourth Edition)*, pages 863 – 913. Academic Press, Boston, fourth edition edition, 2009. ISBN 978-1-59749-272-0. doi: http://dx.doi.org/10.1016/B978-1-59749-272-0.50018-9. URL `http://www.sciencedirect.com/science/article/pii/B9781597492720500189`.

[66] Giang Binh Tran and Eelco Herder. Detecting filter bubbles in ongoing news stories. In *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015), Dublin, Ireland, June 29 - July 3, 2015.*, 2015. URL `http://ceur-ws.org/Vol-1388/latebreaking_paper3.pdf`.

[67] Arjen Van Dalen. Structural bias in cross-national perspective how political systems and journalism cultures influence government dominance in the news. *The International Journal of Press/Politics*, 17(1):32–55, 2012.

[68] Klaus Von Beyme. *Parteien in westlichen Demokratien*, volume 245. Piper, 1982.

[69] David M. Wood. Issue dimensions in a multi-party system: The french national assembly and european unification. *Midwest Journal of Political Science*, 8(3):255–276, 1964. ISSN 00263397. URL `http://www.jstor.org/stable/2108955`.

[70] Hsiang-Fu Yu, Fang-Lan Huang, and Chih-Jen Lin. Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85(1-2):41–75, 2011.