SCHOOL OF COMPUTATION,
INFORMATION AND TECHNOLOGY —
COMPUTER SCIENCE

TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Computer Science

From Hashtags to Ballot Boxes: A Close Look at the 2023 Turkish Election

Efe Sener

SCHOOL OF COMPUTATION,
INFORMATION AND TECHNOLOGY —
COMPUTER SCIENCE

TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Computer Science

From Hashtags to Ballot Boxes: A Close Look at the 2023 Turkish Election

Von Hashtags zu Wahlentscheidungen: Ein umfassender Blick auf die Türkischen Wahlen 2023

Author:	Efe Sener
Supervisor:	Prof. Dr. Georg Groh
Advisor:	Carolin Schuster
Submission Date:	15.03.2023

I confirm that this bachelor’s thesis is my own work and I have documented all sources and material used.

Munich, 15.03.2023 Efe Sener

Acknowledgments

Abstract

Contents

1 Introduction

In recent years, governments and the public have realized the importance of social media, especially Twitter/X, which has a decisive role in mobilizing social and political activism (Uysal & Schroeder, 2019). Twitter has been instrumental in studying human behavior with social media data (Pfeffer et al., 2023), described as a digital social telescope by researchers in the social science field (Mejova et al., 2015). It has provided a somewhat free environment and guided social and political debates to gain new dimensions (Yerlikaya & Toker, 2020), where individual users can directly and publicly address comments to their representatives under conditions of anonymity (Theocharis et al., 2020). The robust rise in Twitter's popularity has stemmed from increasing accessibility to technology and affordability. Millions of people consume news from social media sites like Twitter (Anwar et al., 2021). In Turkey's case, Twitter began to be taken seriously after the unrest in the Middle East, especially after the Gezi Park protests in 2013 (Zaharna & Uysal, 2016), where Twitter was one of the most valuable media for protestor communication, given censorship (Ogan & Varol, 2017). In 2020, Turkey was ranked as the 10th most-used language on Twitter, with around 560 thousand tweets posted daily (Alshaabi et al., 2021).

This thesis aims to analyze the Twitter data, provided by Sabanci University (Najafi et al., 2022), to understand the Turkish Twitter discourse surrounding the May 2023 elections. Using innovative topic modeling techniques, this thesis will discover the most prevalent topics in Turkish Twitter between July 2022 and June 2023. It will uncover how these topics correlate with real-life events and how they reflect the election agendas of parties. This thesis will also compare the results with results observed in other countries. In a non-English-speaking country like Turkey, this thesis furthermore seeks to find solutions to the need for a more thorough and data-driven analysis of political discussions on Turkish Twitter.

In this chapter, the thesis starts by explaining the historical context and then continues to present the current political landscape. It demonstrates the importance of the May 2023 elections, emphasizes the significance of Twitter in Turkish politics, and deep dives into research questions. In the next chapter, the thesis examines various related works, asking similar questions and analyzing their results. After that, the thesis explains the Twitter dataset and used methodologies while collecting and analyzing the data. Next, the thesis deep dives into the analysis results, and later discusses the findings by interpreting them, highlighting both the limitations and future work. The final section summarizes the results and its implications.

1.1 Background

It is crucial to examine Turkey's historical political context to understand the current complex political landscape and the May 2023 elections.

After the collapse of the Ottoman Empire, the Turkish Republic was declared in 1923. Some attempts were made, but the first multi-party elections were held in 1946. Until 1945, the Republican People's Party (CHP) was the only party in the parliament, and until 1950 it was the ruling party. The CHP was founded by Mustafa Kemal Atatürk, also the founder of the Turkish Republic.

With a multi-party system in a young republic, political power was now open to various groups. Different and new ideologies arose and started to organize politically (Rabasa & Larrabee, 2008). The military saw their role as the protector of the Republic and Atatürk's ideologies and overthrew the governments in 1960, 1971, and 1980. The 1980 military coup, which introduced a new constitution, was after a period of political fragmentation and civil instability in the 1970s.

During the 1970s, political Islamism started to emerge, which challenged the secularist nationalism and modernization ideologies of the CHP (Yilmaz & Bashirov, 2018). Changes in the political structure, the constitution, and civil liberties, major economic crises in 1994 and 2001 (Ardan, 2023) contributed to Islamic political groups' political influence and strength, to the emergence of new political players and parties like the Justice and Development Party (AKP) (Rabasa & Larrabee, 2008).

Since 2002, AKP has been in power in Turkey. Out of 15 elections, AKP just lost the local elections in 2019, in which the opposition coalition won more than four significant municipalities. Especially in Istanbul, the opposition won twice because the first election was canceled. For the May 2023 elections, the main opposition coalition was established from CHP, Good Party (IyiP), Felicity Party (SAADET), Democrat Party (DP), and two new parties were established out of AKP: Democracy and Progress Party (DEVA) and Future Party (GP) (Atila, 2022). Even though most of the polls favored the opposition in the May 2023 elections (Saç & Çoban, 2023), AKP has won the majority of the parliament and Recep Tayyip Erdogan was elected in the kickoff elections for the third time as president, after serving two terms as president and two terms as prime minister since 2003.

1.2 Research Questions

This section introduces the research questions guiding this thesis, which are based on qualitative methods to analyze the Twitter discourse surrounding the May 2023 elections in Turkey.

The research questions are divided into two parts. The first part will cover the main research objective of this thesis, which is the analysis of the topic modeling results. The first question is as follows: "What were the most prevalent topics in Turkish Twitter discussions during the May 2023 elections?". This question is necessary to understand the main topics of the May 2023 elections discussed in social media.

The next question is "How do real-life events during the election period correlate with shifts in discussion topics on Twitter, and in what ways do these shifts mirror political movements?". This question focuses on the reflection of real-life events and political movements in Twitter discussions.

The third question is about parties and their election agendas: "How do the Twitter discussions about the ruling party and the opposition during the election lead-up reflect and compare to their respective election agendas and public statements?". This question is essential to understand the reflection of the election agendas of the parties and the differences between them on Twitter.

With these questions in mind, the second part of the research questions covers the comparison of the results of the topic modeling with other research, where a similar approach was used for different countries. The main question is as follows: "How do the key themes, content, and engagement levels in the Turkish Twitter discourse surrounding the May 2023 elections compare with those observed in the past elections in other countries?".

2 Related Work

The recent advances in Natural Language Processing (NLP) and easy access to open-source models allow researchers to study text data by performing sentiment and emotional analysis, topic modeling, semantic search, and many more. Large language models by OpenAI considerably explain how fast the NLP field develops.

In this thesis, topic modeling is performed on massive text data. Topic modeling is an unsupervised tool that helps extract the underlying themes from the given text data. There are several topic modeling approaches, and this thesis focuses on neural topic modeling. Unlike conventional models like Latent Dirichlet Allocation (LDA), a generative probabilistic model introduced by Blei et al. (2003), neural topic models have been used in important NLP tasks, including text generation, document summarisation, and translation, fields to which conventional topic models are complex to apply (Zhao et al., 2021). This thesis uses the neural topic model BERTopic, introduced by Grootendorst (2022), which is explained in detail in Chapter 3.

A tremendous number of studies have applied topic modeling in their research. In the political science field, Ilyas et al. (2020) performed topic modeling using LDA to discover daily discussion topics on Twitter about Brexit and to find out whether the topics discussed on Twitter were representative of actual events taking place, aligning with the second research question of this thesis. They found out that their model was representative of the actual events. Kaiser et al. (2020) used a structural topic model (STM), similar to LDA, to analyze the right media coverage during the 2016 US elections. The analysis shows that a media outlet is identified between the extreme far-right and mainstream right by finding out that they cover extreme and conservative topics. For the 2020 US elections, Anwar et al. (2021) applied topic modeling using BERTopic on pro-Trump tweets to analyze the most mentioned words for each topic and how frequent the topics were, aligning with the first research question of this thesis. Gritto (2022) applied BERTopic along with other German BERT models on Twitter data from German politicians and analyzed their results, aligning with the third research question of this thesis. She discovered that using BERTopic with the Sentence-BERT (SBERT) model yielded more valuable and significant topics. On the other hand, Contreras et al. (2022) used both LDA and BERTopic on Spanish Panamanian parliamentary proceedings. The research suggests that both models perform well with long multilingual political texts despite the small dataset.

It is essential to mention that according to the available literature, few studies apply topic modeling to multilingual political data. For the Turkish language, since the introduction of BERTurk by Schweter (2020), which is based on the BERT model by Devlin et al. (2019) trained on Turkish dataset, the Turkish NLP community is getting bigger and bigger day by day. Recently, a new model called TurkishBERTweet trained on the Turkish Twitter dataset was presented by the same team[1] that released the public social media dataset #Secim2023[2] (Najafi & Varol, 2023). The team has used TurkishBERTweet to conduct daily sentiment analysis and various other analyses on the #Secim2023 dataset, which will be discussed later.

This thesis will build upon the mentioned research and conduct one of the first neural topic modeling researches on a massive political Turkish language dataset using BERTopic. As mentioned in previous research, BERTopic yields more valuable and significant topics than other topic models, which is why this thesis will use that model. Since Najafi and Varol (2023) and also Najafi et al. (2022) analyze the same dataset as this thesis, but with different approaches and questions, this thesis will also use their results while answering the research questions.

[1]Center of Excellence in Data Analytics, Sabanci University, Turkey
[2]Najafi et al., 2022.

3 Experiments

This thesis uses the BERTopic model to apply topic modeling on the #Secim2023 dataset. Before diving into the results and discussion, this chapter explains the dataset, how the tweet hydration[1] is performed on the tweets from the dataset, how BERTopic and neural topic modeling works generally.

3.1 The Dataset

The dataset published by Najafi et al. (2022) consists of tweet IDs collected daily between July 2022 and June 2023, a total of around 250 million tweets. The frequency of the collected tweets is shown in Figure 3.1.

Due to Twitter's Developer Agreement and Policy[2], a public dataset can only include (tweet) IDs. In order to access all tweet information, they must be hydrated. Typically, a year before, a research group would have had access to Twitter Academic API[3] and used packages like Hydrator[4] to gather tweet information quickly. Unfortunately, after Elon Musk bought Twitter, Academic API was restricted and then shut down at the end of May 2023 (Calma, 2023), before the start of this thesis. Today, there are only paid options starting from 100$ for 10,000 tweets per month, 0.3% of what was previously available for free access in a single day.

If one has tweet IDs, other methods exist to hydrate the tweets nowadays. All of the following methods use some embedded retrieval mechanism to gather the tweet information. The first method uses Twitter's official page to retrieve embedded posts or videos given the tweet ID: https://publish.twitter.com. The second method, also used in this thesis, is implemented by React engineers in-house: https://github.com/vercel/react-tweet. One can have a JSON output with sufficient information for analysis by sending HTTP requests and tweet ID as a parameter.

As seen in Figure 3.1, the collected tweets (blue) are less than the total tweets in the dataset (orange). Out of 250 million tweets, only around 150 million tweets are collected. One of the main reasons for that is the deleted tweets. Since the hydration

[1]The process of retrieving a tweet's complete information with only tweet ID.
[2]https://developer.twitter.com/en/developer-terms/agreement-and-policy
[3]https://developer.twitter.com/en/use-cases/do-research/academic-research
[4]https://github.com/DocNow/hydrator

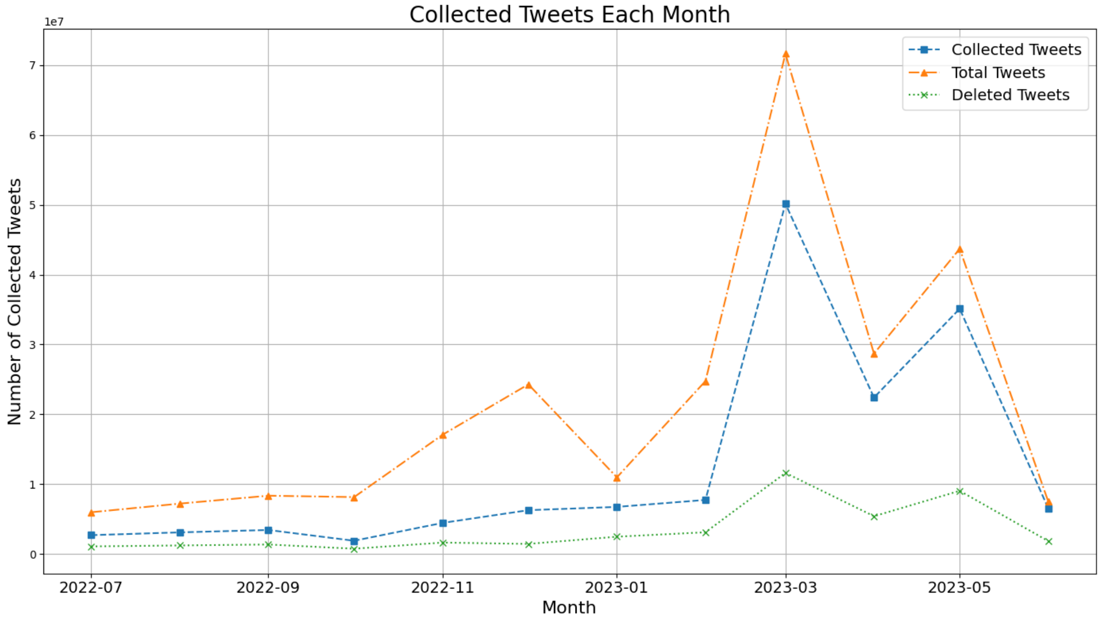

Figure 3.1: The graph displays the monthly count of collected tweets from the #Secim2023 database, spanning from July 2022 to June 2023, in increments of 10 million. The orange line displays the total number of tweets in the database, the blue line displays the total number of collected tweets from the database and the green line displays the deleted tweets on the month of hydration, around December 2023.

timeline for this thesis was between October 2023 and January 2024, and the tweets are between July 2022 and June 2023, there are approximately 50 million deleted tweets in the dataset (green).

There are several reasons why there are lots of deleted tweets. The main reason for the vast number is the deletion of highly interacted original tweet posts. According to Twitter, if the original tweet is deleted, the reposts to that tweet are also not available anymore[5]. The #Secim2023 dataset contains retweets, quotes, and replies in the majority. As discussed by Pfeffer et al. (2023) in their research, almost 80% of all tweets on Twitter refer to other tweets, with original tweets making up the rest.

Furthermore, the question of why people deleted their posts in the first place can be answered in two aspects. First, as mentioned in this New York Times article by Klosowski (2022), whether the person posting is a public figure or not is not essential:

[5]https://help.twitter.com/en/using-x/repost-faqs

companies in the hiring process could run a social media background check, leading to rejection.

Secondly and even more alarming, the government can pull an old tweet out of context and use it against the person, leading to an arrest. The Turkish government's control over social media is widely recognized. It has instituted nationwide bans before and has arrested people accused of "provocative posts" continuously (Scott, 2023). Freedom House's 2023 report states that Turkey's global and internet freedom scores are classified as "not free" (Freedom House, 2023). These reasons could have eventually led to the deletion of many tweets after the election.

The gap between collected plus deleted tweets and total tweets lies under restricted rate limits by Twitter[6]. During hydration, every second or third response was empty, which led to second and third hydration batches of the missing tweets. There are around 50 million tweets that could not be collected. Due to time limitations and the lengthy duration of big data analysis, the hydration process resulted in the maximum feasible collection of tweets within the constraints.

3.2 The Methodology

As mentioned in the previous chapter, topic modeling is an unsupervised tool that helps extract the underlying themes from the given text data. BERTopic is a neural topic model, one of many topic modeling approaches. Due to time constraints and the time plan of this thesis, only the BERTopic model is used for topic modeling. Since several methods could be used, it is important to mention why BERTopic is used and why the others are not. Egger and Yu (2022) found out that for short and unstructured texts like Twitter data, BERTopic can extract contextual information, and it offers the most potential compared to different embedding-based topic models like Top2Vec. According to their research, BERTopic has high versatility and stability across domains and supports different topic modeling variations. Like other embedding-based topic models, it allows multilingual analysis, and there is no need for preprocessing of the original data. However, the embedding approach might cause too many topics and outliers in some cases, which makes the results more challenging to interpret and should be examined in detail. Some long documents could occasionally involve multiple topics, but in this approach, every document is assigned to a single topic, which could be a disadvantage.

[6]https://business.twitter.com/en/blog/update-on-twitters-limited-usage.html

3.2.1 The BERTopic Pipeline

According to Grootendorst (2022), BERTopic generates topic representations in six steps, shown in Figure 3.2. First, without preprocessing, each document must be embedded using a pre-trained model. In this thesis, the SBERT model is used, introduced by Reimers and Gurevych (2019). SBERT modifies the BERT model and derives semantically meaningful sentence embeddings, also from multilingual documents, allowing tasks like clustering or information retrieval via semantic search. SBERT also allows the selection of various pre-trained multilingual models supporting more than 50 languages[7]. This thesis uses the `paraphrase-multilingual-MiniLM-L12-v2` model, which supports Turkish and is the fastest and one of the best performers among other multilingual models (Reimers & Gurevych, 2020). This part of the pipeline allows to do chunk embeddings, saving the results and using them later, making the big data analysis easier with restricted hardware availabilities.

Secondly, the dimensionality of these resulting embeddings is reduced to optimize the clustering process by the Uniform Manifold Approximation and Projection (UMAP) algorithm, which plays a massive role in big data analysis (McInnes et al., 2020). Afterward, these low-dimension embeddings are clustered using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) technique, and the resulting clusters consist of semantically similar documents. This step allows unrelated documents to be assigned as noise or outliers, which improves the result, and both of these steps can be influenced very much by changing the parameters of the algorithms.

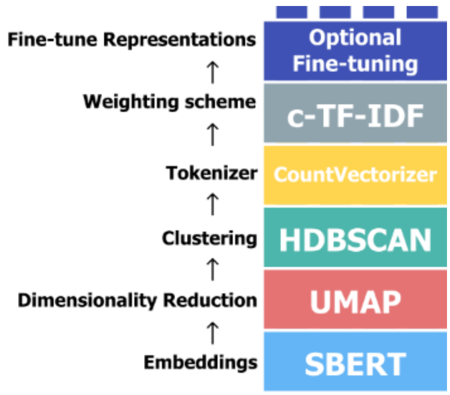

Figure 3.2: Default BERTopic Algorithm[8]

Fourthly, each cluster is tokenized using a Vectorizer like CountVectorizer[9]. Together with the weighting of these tokens, where a custom class-based variation of the Term Frequency – Inverse Document Frequency (c-TF-IDF) algorithm is used, they are responsible for creating the topic representations. Like the previous step, this step also allows room to play with various parameters to tune the model, which affects the results considerably.

At last comes the topic representation, where the topics can be fine-tuned using

[7]https://www.sbert.net/docs/pretrained_models.html
[8]https://maartengr.github.io/BERTopic/algorithm/algorithm.html
[9]https://maartengr.github.io/BERTopic/getting_started/vectorizers/vectorizers.html

various methods. This thesis uses KeyBERTInspired and Maximal Marginal Relevance (MMR) models[10], which can be easily imported from the BERTopic library. They leverage the c-TF-IDF algorithm and weights keywords to represent the related topics. This thesis also leverages the power of LLMs by OpenAI, specifically GPT-4 Turbo, to better represent the resulting topics. One can also leverage other open-source LLMs, but most only support English or a few other languages.

3.2.2 The Analysis Srategy

For hydration, this thesis used the *Social5* server provided by Research Group Social Computing[11], which had sufficient hardware capabilities to run several scripts in parallel, and each of them was able to leverage multithreading, speeding up the process.

The result of the hydration was around 40 GB, which makes the analysis in one batch impossible. For this reason, hydrated tweet information was divided into smaller batches to make the analysis more manageable later. Before the topic modeling analysis, the resulting data is cleaned by extracting relevant information. That means only extracted tweets that include some text. These are then converted into lists and pulled into the analysis environment.

In the next step, to speed up the big data analysis process, this thesis used Google Colab[12] Pro that brings A100, V100, or T4 GPUs to usage, making the process considerably faster.

As mentioned before, because the data is big, it does not fit into the BERTopic model in one batch due to hardware restrictions. To tackle that issue, 2.5 million tweets are randomly sampled in proportion to the total number of collected tweets monthly, which makes up 1% of the total number of tweets in the #Secim2023 dataset.

The strategy used in this thesis was to first train the model with the sample tweets to find relevant topics in Turkish political discourse on Twitter. After that, all the remaining tweets were assigned a topic using the trained model without any need to train an additional model. This strategy also allows the analysis to be done in batches to make it smoother and more manageable.

The hydration, preprocessing, and analysis scripts mentioned are open-source and can be accessed on GitHub[13]. The analysis of this thesis is mainly based on two notebooks written by Maarten Grootendorst, father of BERTopic. The first is called

[10]https://maartengr.github.io/BERTopic/getting_started/representation/representation.html
[11]https://www.soc.cit.tum.de
[12]https://colab.google
[13]https://github.com/EfeSenerr/Thesis

"Best Practices" and aims to get the best results[14]. In contrast, the second one is called "Big Data" and focuses on analyzing big data efficiently by optimizing and leveraging the power of the GPU[15]. The parameters used in the models are also inspired by them, resulting in minimal changes to optimize the results.

The changes also followed Grootendorst's recommendations[16]. However, while making minimal changes in the parameters, the results can be highly affected by these changes. In this thesis, two approaches are made, resulting in two relatively different results while being similar in most topics. The results of this process are presented and discussed thoroughly in the following chapters.

In short, the following parameters changed from their default value are discussed.

In UMAP, `n_neighbors` is set to 25 from the default 15 to obtain a more global view of the embedding structure, resulting in larger clusters. Other parameters in UMAP followed the default values.

The `min_cluster_size` parameter in HDBSCAN is among the most critical parameters. It is set to 100 from 10 as a default to increase the minimum size of a cluster, thereby reducing the number of clusters the model generates. In the first analysis, this parameter was set to 20, and after getting too many topics, the parameter was increased to 100. For a massive dataset like #Secim2023, it could make sense to increase this parameter even more, but due to time restrictions, this thesis could not test this additionally. The `min_samples` parameter is usually automatically set to the value of `min_cluster_size`, but in this case, it is left in 20 to reduce the number of outliers generated. In the second analysis, the `prediction_data` parameter is set to `True` to predict new points later.

For passing out to CountVecorizer while training the model, one can prepare the vocabulary of the dataset used in training beforehand so that the tokenizer does not need to do the calculations later. It also reduces the necessary RAM during the training. This thesis creates a vocabulary of words and parses them such that they need to appear at least ten times in the data while removing the stopwords using Turkish stopwords from the NLTK library.

KeyBERT and MMR are used as the representation models of the topics. GPT-4 Turbo is later used for only the top 50 topics to reduce the costs constructed from it.

All these parameters mentioned above are passed to BERTopic along with its sampled text data and pre-computed embeddings. One can also play with some parameters of the BERTopic, but this thesis did not use any additional parameters. Some parameters do not play a critical role. For instance, the `min_topic_size` parameter in BERTopic is automatically set to the value of `min_cluster_size`. After several hours, this process

[14]https://maartengr.github.io/BERTopic/getting_started/best_practices/best_practices.html
[15]https://huggingface.co/MaartenGr/BERTopic_Wikipedia
[16]https://maartengr.github.io/BERTopic/getting_started/parameter%20tuning/parametertuning.html

resulted in topics that are the baseline of all topics in Turkish political discourse during the May 2023 elections. The resulting collected tweet data is then, in sequence, embedded and "transformed" by the trained topic model, which assigns topics to individual tweets. These results are demonstrated in the following chapter.

4 Results

As explained in the previous chapter, this thesis follows several topic modeling approaches to obtain the most representative result that can be analyzed. To emphasize the approaches again, the first approach follows Martin's general recommendations and does not change any arguments. This BERTopic approach delivers around 7500 topics, which is too much to analyze and make sense of the results.

The main reason behind the mass number of topics is `min_cluster_size` remaining at 20, allowing around five thousand topics with less than five thousand tweets. The other main reason is the diversity of the analyzed data. To determine the topics, sample data of 1% of all collected tweets is used from a dataset that contains more than 200,000 daily tweets. When manually looking at the data, one can realize that even though the size is small, the topic can be related to a real-life event. Nevertheless, on the other hand, most of the topics can be combined or be in the outlier category.

With the experiences from the first approach, the second approach tries to minimize the cluster size of the topics to obtain a better result. When the `min_cluster_size` remained at 100, around 1000 topics were returned from the topic model. Similar to the first approach, this approach also covered the most significant topics. Because the cluster sizes are bigger with this approach, the topics are more challenging to label and categorize in one area. Although one can realize the similarities between these approaches in the top eight representative words, the representative tweets differ. These representative tweets are used in OpenAI prompt to label the topics, so the quality of these tweets is of high importance. In this approach, most representative tweets include three similar long texts with a high proportion of hashtags and mentions. That is also why almost half of the top 20 topics are demands or requests from the government, where many users and bots spam almost similar tweets to put the topic in trends, promoting politicians, parties, and specific agendas (Najafi et al., 2022).

After considering these arguments, this thesis mainly analyzes the results of the first approach, where one can realize topic trends consisting of specific topics that make comparing and analyzing the results considerably more straightforward.

It is essential to mention that outliers are not mentioned via graphs but are remembered during the results. Although some approaches in this tried to redistribute outliers, outliers exist for a reason. As Grootendorst notes, outliers are to be expected, and pushing the output to have no outliers may not correctly represent the data.

One disadvantage of following the default pipeline of BERTopic is that when the model is first trained, the biggest cluster is usually the outliers. However, when training the model and transforming the rest of the data, the outlier cluster is considered a usual topic cluster like any other. For instance, during one of the approaches, the model had more tweets in the outlier cluster while training than after transforming all the collected data, which might, on the one hand, cause lots of meaningless topics and, on the other hand, not accurately represent the data. There are several new approaches that BERTopic supports that overcome that issue. One can try *Online Topic Modeling* to incrementally train the model, or *Merge Multiple Fitted Models* approach, where multiple models are trained and, in the end, merged so that no information is lost. These approaches are discussed in Section 5.4.

The top results of the topic modeling are presented in Table 4.1, ordered by the number of tweets in the clusters. The Table 4.1 shows the topic labels on the left and their respective topic's eight representative words on the right. Although it would be hard to translate the representative words, the labels are labeled and then translated by GPT-4 Turbo for better interpretation. One can look at the GitHub repository[1] for the original results of the model for the top 50 topics. Since OpenAI models are not cost-free, they have been only used for the top 50 topics.

The top 15 topics represent various themes around Turkish political discourse. These topics include general political discussions, praise and criticism of both ruling and opposition parties, wishes and demands from the government, and the earthquake disaster.

Before diving deeply into the topics and understanding the big picture, clarifying some topics for better context would be beneficial. Topics like *Political Discussions*, *Political Discussions & Critism*, *Critism of Political Figures*,

Praise of Political Figures are more general and include tweets with positive and negative sentiments towards both the ruling and the opposition parties. One can realize that the representative words of the first topic include Twitter user names, which are active accounts that share news mainly about the elections. A high number of replies and a high level of interaction with their tweets made this category the most noteworthy.

However, topics like *Opposition Candidates*, *Opposition Coalition & Criticism*, *Opposition Figures & Preferences* and *Erdoğan's Election Chance* focus on and express opinions about specific target groups, either the ruling party or the opposition parties and their candidates.

It is hard not to notice the fourth biggest cluster *Religious Wishes*, which includes tweets that cover aspects of prayer and Islam. While most of the representative sentences

[1] https://github.com/EfeSenerr/Thesis_paper

Topic Label	Representative Words
Political Discussions	aysedogan1955, yirmiucderece, hassa61, furkancerkes, secimtr2023, cenginyurt52, aykiricomtr, pusholder
Opposition Candidates	oy, kılıçdaroğlu, oyları, seçmeni, oylar, vermem, aday, seçimde
Erdoğan & Nationalism	türk, turkey, türkiye, yüzyılı, türkiyeyüzyılı, stanbul, milliyetçisi, cumhuriyeti
Religious Wishes	allah, versin, eylesin, müslüman, razı, mekanı, namaz, dini
Political Discussions & Criticisms	cır, terketmek, denilince, uyruklu, ayrilmis, muharremin, vezir_yuce, bahtiyar_ergn
Criticism of Political Figures	siyaset, siyasetin, siyasetiniz, siyasetçi, siyaseti, batsın, siyasete, siyasette
Opposition Coalition & Criticism	erbakancayazan, örneğinde, yolunuzdan, iddialar, atıldılar, seçicez, ihtimalde, omurgasızlıktır
Treason Accusations	ülkeyi, ülkeye, ülke, ülkenin, ülkede, yağdanlıkları, unutturacaksınız, sığınmayın
Opposition Figures & Preferences	eu, me, que, aq, pra, amk, tudo, dedem
Retirement System	emeklilik, kısmi, kademeli, emekli, emeklilikte, prim, 5000
Earthquake & Demands	neticesiz, uzunca, deprem, depremin, süredir, atanma, yumuşak, depremde
Provocation & Discussions	kızmaz, ahhh, vura, yanarız, soğancı, mitink, arsibjk1903bjk, baskan
Praise of Political Figures	başkanım, başkan, başbakan, başkanın, cumhurbaşkanım, selo, mehmetfatihser5, mehmetersoy57
Teachers & Demands	öğretmen, öğretmenler, ataması, ilave, ücretli, öğretmenlere, cumhuriyetimizin, öğretmenlerin
Erdoğan's Election Chance	rdoğana, erdoğan, erdoğanı, egemenlere, turda, tura, vereceğim, oyumu

Table 4.1: Result of the topic modeling with outlier redistribution. The top eight representative words from MMR model for each of the top 15 topics with their respective topic label translated into English.

show tweets expressing condolences to various political figures who have lost their relatives, many of these tweets also include phrases of support and sympathy, such as *'God bless you'*, underlining solidarity.

The label *Treason Accusations* might sound absurd. The main reason for that is, again, the representative tweets that include some accusations against the political figures, blaming them with extremes like committing treason or being a terrorist. As one of the largest clusters, topics like this illustrate the polarization within the Turkish political landscape, which has been prominent over the years (Çevik, 2018). Another topic in this table, *Provocation & Discussions*, also highlights the polarization level in the Turkish

Twitter political discourse. The representative tweets are aggressive and show a high level of emotions.

Erdoğan & Nationalism topic and its representative words show a significant degree of nationalist ideas. The label itself includes Erdoğan because of the representative tweets mentioning Erdoğan and the tenth representative word *Erdoğan*, which is not seen in the table. However, it is essential to mention that both the opposition and ruling party coalition include parties based on Turkish nationalism that played a significant role in May 2023 elections, parties being IyiP and Nationalist Movement Party (MHP).

Although it is more explicit in other topic modeling strategies that this thesis tried, one can also see the volume of topics like *Teachers & Demands* and *Earthquake & Demands* in Table 4.1. The role of Twitter in the latter topic is crucial. Because of the earthquake on 6 February, all controversial telecommunication lines were disrupted, where Twitter has been widely used for rescue operations (Çevik & Aksoy, 2023).

All these mentioned topics are deeply analyzed in the following section, with graphs supporting understanding of the topics and their trends.

4.1 Detailed Analysis of Topic Categories

For simplicity and better understanding purposes, this thesis analyzes the topic modeling results in three categories: topics related to the ruling party visualized in Figure 4.1, the opposition visualized in Figure 4.2, and the remaining topics visualized in Figure 4.3.

These graphs visualize the selected topics in a normalized monthly distribution. Normalization has been done considering the top 50 topics. Otherwise, the interpretation of topics and their trends was almost impossible. One can not realize any trends in a frequency graph because the #Secim2023 dataset is not collected equally between months, demonstrated in Figure 3.1.

In all these graphs, the biggest cluster *Political Discussions* is shown as the baseline in a straight blue line to compare these different graphs and analyze the trends more smoothly.

4.1.1 Ruling Party Discourse Analysis

Beginning the analysis with the first graph 4.1, it covers the monthly distribution of the ruling party and Erdoğan-related topics. It is essential to acknowledge the baseline topic *Political Discussions*, keeping its importance in the months to elections, massively increasing after the start of 2023 up to more than 10% of the top 50 tweets. That is normal, considering that in the months leading up to the elections, almost all the news was political, and rallies around Turkey were happening daily. Undecided opposition

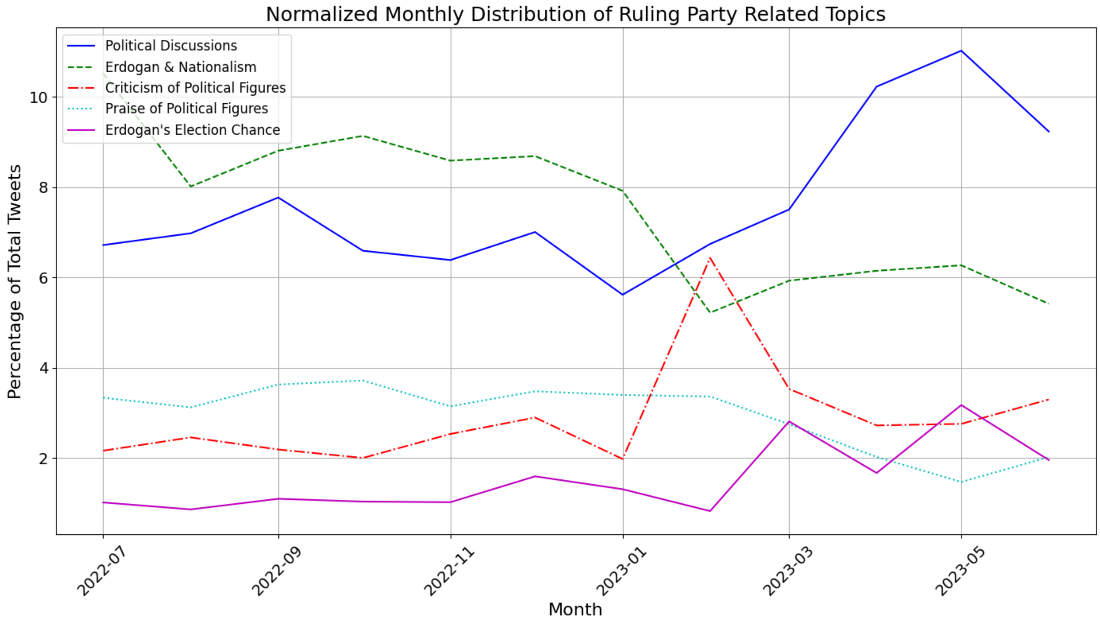

Figure 4.1: Percentage of ruling party related topics in the top 50 topics. The blue line represents the biggest cluster and is the baseline, and the other lines represent topics related to Erdoğan and the ruling party.

candidacy, hot topics like the economic crisis and migration policy, and uncertainties in domestic and foreign policies led social media to a political discussion hub.

Having the baseline topic covered, the next topic with a green dotted line covers the topic *Erdoğan & Nationalism*. As mentioned previously, it is essential not to converge Turkish Nationalist ideas with Erdoğan and the AKP regime. There are nationalism-based parties in both the opposition and the ruling coalition. The opposition coalition's name is even 'Nation Alliance', in other words, the 'Table of Six'.

Turkish Nationalism has always been one of the base ideas of Turkish politics since the republic's foundation. The reason for that is now the opposition and once the founding party, CHP, who laid Turkey's six founding fundamental ideologies called the 'Six Arrows'. One of the *arrows* represents Nationalism. However, as Hotelling's model suggests, throughout the years, because CHP's idealogy moved more toward the left wing, parties like AKP or newly established parties like ZP started to fill the empty right-wing spots (Caramani, 2020). Representative tweets showing support for different political figures using nationalist ideas also underline this aspect.

The migration policy was one of the hottest topics of the May 2023 elections. The government's attitude of welcoming more than six million refugees while getting funding from the European Union for this purpose sparked lots of tensions amongst different ideologies and parties. In 2021, a new party called the Victory Party (ZP) was formed from the ranks of the opposition. It then rapidly increased its popularity as a mainly single-issue party focused on expelling refugees back to their countries (Esen, 2022a). ZP formed the second opposition coalition against Erdoğan, and their candidate received more than 5% of the votes.

These events highlight why Nationalism has been one of the most trending topics during the months up to the election. However, in the new year, more specifically after the 6 February earthquake, other topics gained more significance, which can also be seen in the Figure 4.1.

One of the topics that gained relatively more importance is the *Criticism of Political Figures*, visualized as a red dotted and straight line in Figure 4.1. On the other hand, the topic *Praise of Political Figures*, visualized as a turquoise dotted line, has lost volume in time. The main explanation for that is the 6 February earthquake, which immediately changed the course of the elections by changing everybody's attention towards the southeast of Turkey.

The following reasons, also underlined by Çevik and Aksoy (2023), can explain the change in the trends of these topics, where criticism outweighed the praise of political figures on Twitter. The system lacked law enforcement until the earthquake, where the contractors of the building constructions sought to maximize their profits without taking the necessary precautions. The weakened state capacity led to the slowness of emergency response during the catastrophe. The military was missing, and there were no plans against this kind of emergency, or in other words, if there was a plan, it was not executed thoroughly. NGOs and volunteers organized themselves and worked together for days after the earthquake. Twitter was always the central platform for organizations and cry for help. Nonetheless, the government slowed down Twitter to prevent 'disinformation' from spreading, which worsened this process.

The last topic in this graph is *Erdoğan's Election Chance*, visualized as a pink straight line. It can be seen that this topic remained relatively low and reached the bottom rock during the month of the earthquake, and in March, it increased a lot and outranked the other two topics in the election month. The reasons for this increase will be discussed thoroughly in Figure 4.2. However, in short, the lack of collaboration and not being decisive about a candidate among the opposition ranks make this increase feasible. Polls also support this argument (Çevik & Aksoy, 2023).

4.1.2 Opposition Discourse Analysis

The analysis continues with the other side of the table, Figure 4.2, which covers the monthly distribution of the opposition-related topics. As a reminder, *Political Discussions*

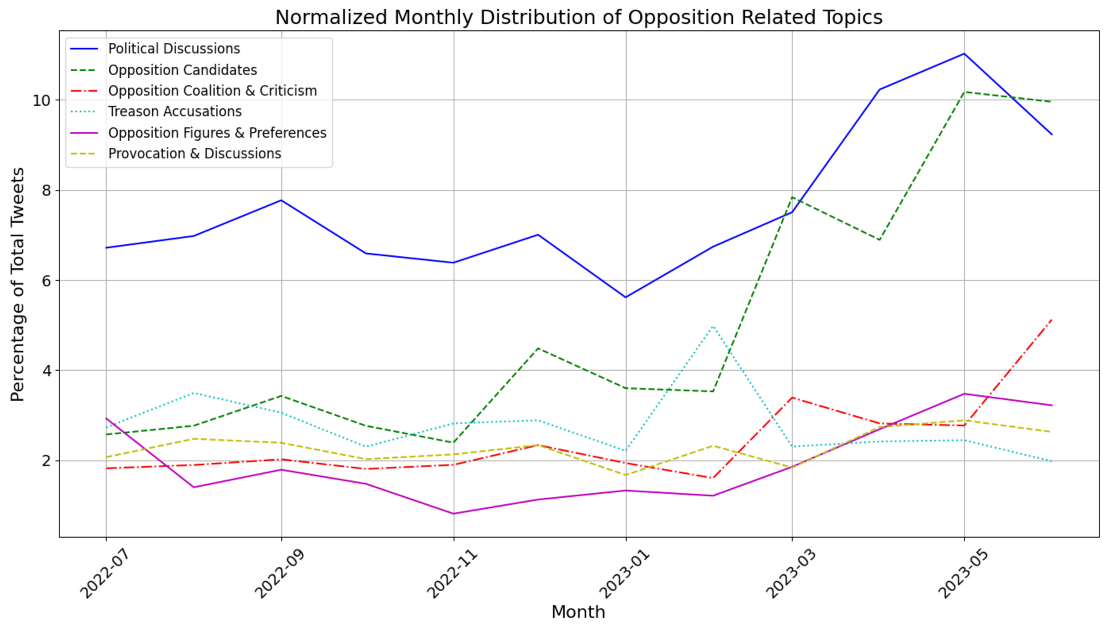

Figure 4.2: Percentage of opposition related topics in the top 50 topics. The blue line represents the biggest cluster and is the baseline, and the other lines represent topics related to opposition parties and their candidates.

topic can also be found in this graph, which is used as a baseline for comparing the graphs.

Continuing with the second topic, *Opposition Candidates* visualized as a green line in the graph. While the representative words cover topics like vote, the main opposition coalition Nation Alliance's candidate Kılıçdaroğlu, and elections, the representative tweets are about people's opinions about who they would vote for in the opposition ranks. For instance, some of them state that *'they would not vote for Kılıçdaroğlu, who does not have any victory against Erdoğan for all these years.'*, or *'they would only vote for Mansur Yavaş'*, who is the mayor of the capital city Ankara.

It is essential to present some context here. The main opposition coalition, Nation Alliance, a.k.a. 'Table of Six', consists of six parties that signed a joint manifesto in 2022 outlining plans to abolish the presidential system and restore the rule of law and civil liberties by returning to the previous parliamentary system (Esen, 2022b). The coalition wanted to achieve this goal by selecting a joint presidential candidate. Like most people in the opposition ranks, (Esen, 2022b) saw selecting a joint presidential candidate as the most demanding task. Since the president holds immense powers in the Turkish presidential election system, the parties were uncertain whether to hand that authority to one individual, so they sought to concentrate more on the alliance than the candidate. Due to these reasons, the coalition pushed this matter further back and could only start to discuss these matters after the earthquake, which is easily remarkable on Figure 4.2.

In March 2023, the majority of parties in the coalition then agreed on Kılıçdaroğlu, who was the leader of the main opposition party CHP and also played an essential role in the establishment of this coalition. After the agreement, the second largest opposition party IyiP also considered that although Kılıçdaroğlu's collaborative personality and focus on unity was an asset for a democratic transition, but stood behind the idea that this could weaken his electoral prospects against Erdoğan. As noteworthy evidence, one could show the most recent polls showing that he follows the popular mayors of Ankara, Mansur Yavaş; Istanbul, Ekrem Imamoğlu, and even the IyiP leader Meral Akşener (Esen, 2022b).

The trending level of the topic also aligns with the social movements at that time (Gültekin, 2023). Just before the earthquake, on 5 February, Kılıçdaroğlu's candidacy topic sparked again. There were campaigns against Kılıçdaroğlu on the streets and also on social media with a slogan *'Do not be a candidate Kılıçdaroğlu.'*. People were holding pan cards with the same slogan in front of the CHP general headquarters and sharing it on Twitter (Ulaş, 2023). Some influencers even planned to march to the headquarters the next day to prevent his candidacy. That is one of the main reasons why the earthquake on 6 February changed and deeply affected the discourse of the May 2023 elections.

IyiP considered Kılıçdaroğlu lacked sufficient public appeal to win the elections. That is why Akşener opposed his candidacy and supported the mayors of Ankara and Istanbul. In short, after several days, the parties agreed on the vice presidency of these two mayors and continued the election campaign as a united front. All these events resulted in the topic *Opposition Candidates* naming itself as the most articulated topic in March on Twitter, surpassing the baseline. Even after the election month, the topic continued to be relevant, beating the baseline again. The discussion rallied then around whether Kılıçdaroğlu was *'the right candidate'*. Several months after the elections, Kılıçdaroğlu lost the in-party elections after leading the main opposition party for 13 years (Gündoğan et al., 2023).

The next trending topic, *Opposition Coalition & Criticism* visualized as a red straight and dotted line, follows a similar tendency, gaining importance in March and continuing to grow even after the elections. However, this topic mainly focuses on direct criticism against the opposition parties. For example, the representative tweets in this cluster complain about Ali Babacan, the leader of DEVA, called him a puppet whose' congressman candidates were entering the elections in the same list as CHP. On the other hand, some tweets refer to Kılıçdaroğlu as a traitor to his own party for this very same reason.

The following topic, *Treason Accusations* visualized as a turquoise dotted line, shows accusations against those who did not support the same idealogy as the writer of these Tweets. Blaming political figures with extremes like committing treason or being a terrorist is a standard tongue that brings the polarization in the Turkish political discourse to light. One can observe that this topic gained importance in February. The reason can be easily tied to the earthquake, such as people accusing the government of lacking law enforcement or not taking the necessary measures for the earthquake.

Represented as a pink straight line, *Opposition Figures & Preferences* highlights a mixture of tweets that include swear words, foreign words and sentences, and discussions around opposition figures. One can understand the abbreviations of Turkish swearwords in the representative words column in Figure 4.2, and it is likely that the BERTopic topic model could not make any sense of the words and merged the topic with foreign words and sentences. Another point the topic model had difficulties understanding is the word *'dedem'*, translated to *'my grandfather'*. Grandfather nickname has been used by the younger generation on social media and overall to refer to the opposition candidate Kılıçdaroğlu. The nickname itself has more of a positive sentiment, emphasizing his kindness and, on the other hand, his age. Although this topic beat all other topics than the baseline topic in July 2022, it started to gain importance when the elections were getting closer and peaked during election month, possibly related to unexpected election results, including comments about the opposition.

The topic, *Provocation & Discussions*, marked by a yellow dotted line, indicates the polarization level in the Turkish Twitter political discourse, as stated before. The representative tweets are aggressive and show negative sentiments, with some of the tweets accusing political figures of being provocateurs. The trend of the topic continues steadily, increasing slightly just before the election month.

4.1.3 Other Relevant Discourse Analysis

The analysis proceeds with the last graph, as presented in Figure 4.3, showcasing the monthly trends of other than ruling party or opposition-related topics. As the baseline, *Political Discussions* topic can be seen as the blue straight line in the graph. One can realize that multiple topics have crossed the baseline topic over the months up to the election.

Kicking things off with the topic *Religious Wishes*, visualized as a green line, the topic contains tweets that cover aspects of prayer and Islam. It features representative words like *'Allah'*, *'müslüman'*, and *'din'*, which than translates to *'Muslim'* and *'religion'*. Some representative sentences exhibit tweets praising political figures from different ranks, saying *'God bless you'*. Although the topic remained at the same level as the baseline and peaked during the month of the earthquake, afterward, it kept its importance while

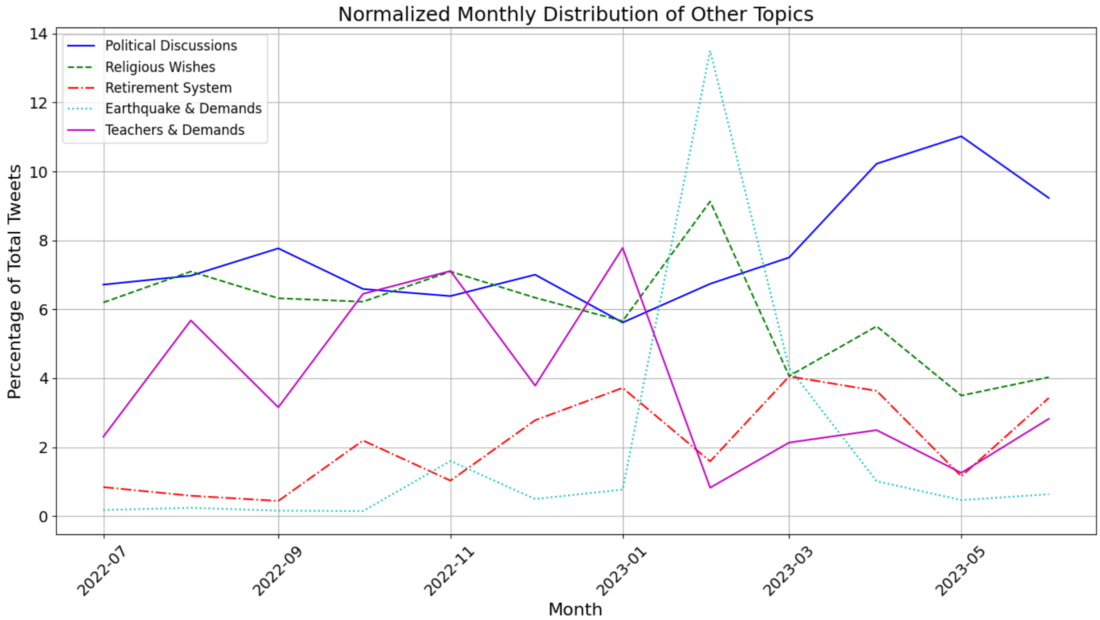

Figure 4.3: Percentage of other related topics in the top 50 topics. The blue line represents the biggest cluster and is the baseline, and the other lines represent topics like demands from the government and the earthquake.

falling in volume, leaving room for political discussions.

Illustrated with a red straight and dotted line, the *Retirement System* topic demonstrates tweets around the retirement system policies. Before the elections, one of the government's promises was to renew the last social security legislation in 1999. At the beginning of March, the government passed new legislation that softens existing rules on retirement age and retirement eligibility requirements, allowing almost more than 2 million workers to retire immediately (Erdem & Bişgin, 2023).

With the enactment of new legislation and a massive increase in the minimum wage, the government increased its spending massively, which resulted in increased public opinion just before the elections (Çevik & Aksoy, 2023). One can recognize that the topic peaked in the month of enactment, slightly lowering in volume afterward.

The topic *Earthquake & Demands* is self-explanatory, highlighting the earthquake on 6 February. In the earthquake, more than 53,000 people lost their lives, which created a social trauma and changed the course of the elections. Although the government mishandled the earthquake, as Çevik et al. (2023) shares, the opposition failed to get the votes from the region, and Erdoğan maintained his popularity in the region, where he received more than 70% the votes in some provinces (Michaelson & Narlı, 2023). As mentioned previously, the role of social media in the earthquake is noteworthy. It is remarkable that the topic peaks massively in February. However, it is very interesting, and maybe also not expected, that it lost almost all of its volume in just two months, making just around 1% of the tweets afterward.

The last topic in the examination is *Teachers & Demands*. The representative sentences exist of sentences like *'We demand 100 thousand additional teacher positions.'*. One of the representative words *'cumhuriyet'*, translated to *'republic'*, points out that tweets exist of a rhyme, calling for an additional 100 thousand teachers at the 100. anniversary of the Turkish Republic. The massive volume of this topic can not be unseen, almost beating the baseline topic two times. It is essential to mention that, as Pfeffer et al. (2023) also points out, Twitter has been used through bots and different means of coordinated automated activities to show influence and power in the public discourse by making the relevant topic trending and keeping it before eyes.

Regardless, this topic aligns with the agenda of the Nation Alliance, where the coalition promised to create new positions for the teachers in one of their rallies (Haber, 2023).

4.2 Supplementary Visual Insights

This chapter presents additional visual representations of the topic model results.

The first visualization is inspired by Maarten Grootendorst, who visualized the topic model results from millions of Wikipedia pages in his Huggingface repository[2]. Like this thesis, he also used BERTopic for big data analysis and visualized the results in the end.

The Figure 4.4 presents the top 50 topics, which have been reduced to two-dimensional embedding space using UMAP. It pictures different clusters with different colors with respect to their sizes and does not contain any outliers. The white background could not be seen if the figure would have included outliers. The topics are illustrated by their top three representative words. One can realize that their top three representative words can differ from the ones from the previous table, Table 4.1.

This is because a different BERTopic pipeline has been used for this visualization. This pipeline includes pre-computing of required parameters of BERTopic, which are generally computed during the fitting procedure. Different from the training pipeline of the base topic model explained in Subsection 3.2.1, the first step starts with UMAP, which has a parameter `n_components` of two instead of five. That leverages the power of GPU by using the cuML library and reduces the topics to two-dimensional space.

[2]https://huggingface.co/MaartenGr/BERTopic_Wikipedia
[3]https://huggingface.co/MaartenGr/BERTopic_Wikipedia

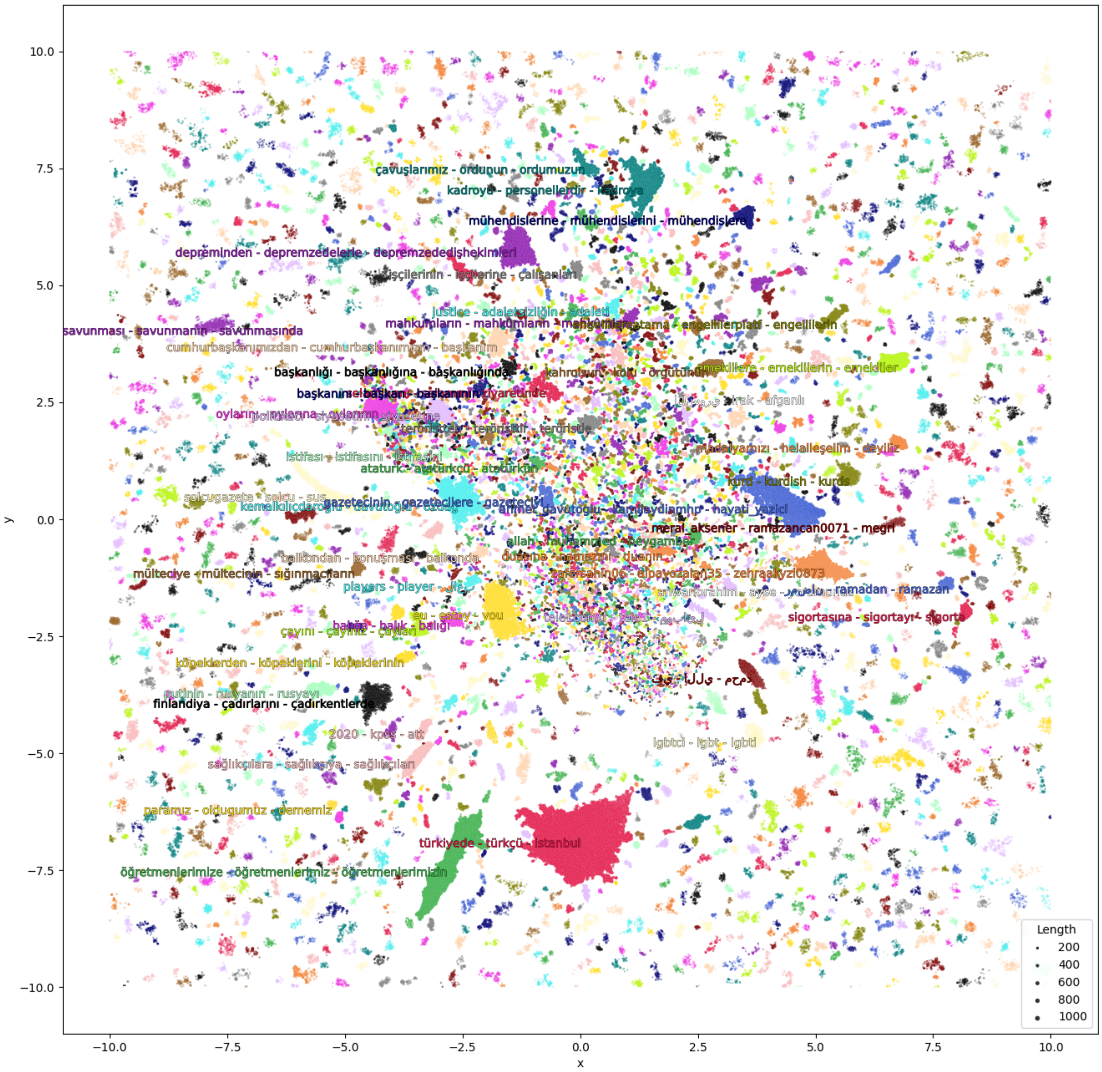

Figure 4.4: The top 50 topics visualized and reduced to 2-dimensional space using cuML's UMAP, following Maarten Grootendorst's guide.[3]

Afterward, HDBSCAN uses the two-dimensional reduced embeddings to pre-compute the clusters of the topics from scratch. These results can be fed to BERTopic to perform manual topic modeling[4]. If the model already has topic labels, it can use those to label them using various representation models.

With embeddings, reduced dimensionality and clusters on hand allow the speeding

[4]https://maartengr.github.io/BERTopic/getting_started/manual/manual.html

up of the fitting process. Additionally, Grootendorst presents a trick to get better results. He states that if one directly gives BERTopic the reduced embeddings, BERTopic would use those to create topic vectors, which could be better. For better results, he suggests giving the model the full embeddings and creating a custom dimensionality reduction class that will return the reduced embeddings. Further details can be found in the thesis GitHub repository.

The results of this pipeline are then used for visualizing the Figure 4.4, with the usage of `seaborn` and `matplotlib` library.

Some of the topics are relatively straightforward to pick up. The enormous red cluster on the bottom of the figure, titled as *'türkiyede – türkçü – istanbul'*, visualizes the topic GPT-4 Turbo labeled as *Erdoğan & Nationalism* from the default pipeline. Next to that topic to the left, visualized as a green pipeline, is the topic *Teachers & Demands*.

The middle of this figure is considerably chaotic. Almost all of the clusters in the middle are related to the election. The topics from the political figures, votes, major Twitter accounts, and the resignation of known figures are covered. At the top of the figure, the cluster related to the earthquake can be seen as purple, titled as *'depreminden – depremzedelerle – depremzede'*.

The following visualization, Figure 4.5, is about the hierarchical clustering of the top 50 topics of the sampled data and has been created using BERTopic function `visualize_hierarchy()`. The fitted BERTopic model has been used, and the parameter `top_n_topics` has been set to 50, which returns that figure. A ward linkage function performs the hierarchical clustering based on the cosine distance matrix between topic embeddings. Since the GPT-4 Turbo labeling has been done after fitting the model, and only for the top 50 topics, the labels in the figure show default labels after the fitting. One can look at the English labels of each top 50 topic in the GitHub repository of the thesis.

If the Figure 4.5 is deeply analyzed, the topics are divided into different colors, which are then connected together to build the total representation. The colors present 13 main subjects among the top 50 topic representations.

Beginning the exploration with the top topics, the green cluster represents the demands of military sergeants willing to be permanently employed. A batch of sergeants who worked on a limited contract in the state institutions for almost 36 years sought permanent employment.

The subsequent topic, visualized as red, addresses the demands of young people endorsing their internships to be recognized as the start date for their insurance, a critical factor determining the start of retirement eligibility.

The following three topics, visualized as turquoise, brown, and green, represent the concerns of agriculture engineers and their work appointments, the retirement system, teachers, and their demands, respectively. The last topic in the top cluster

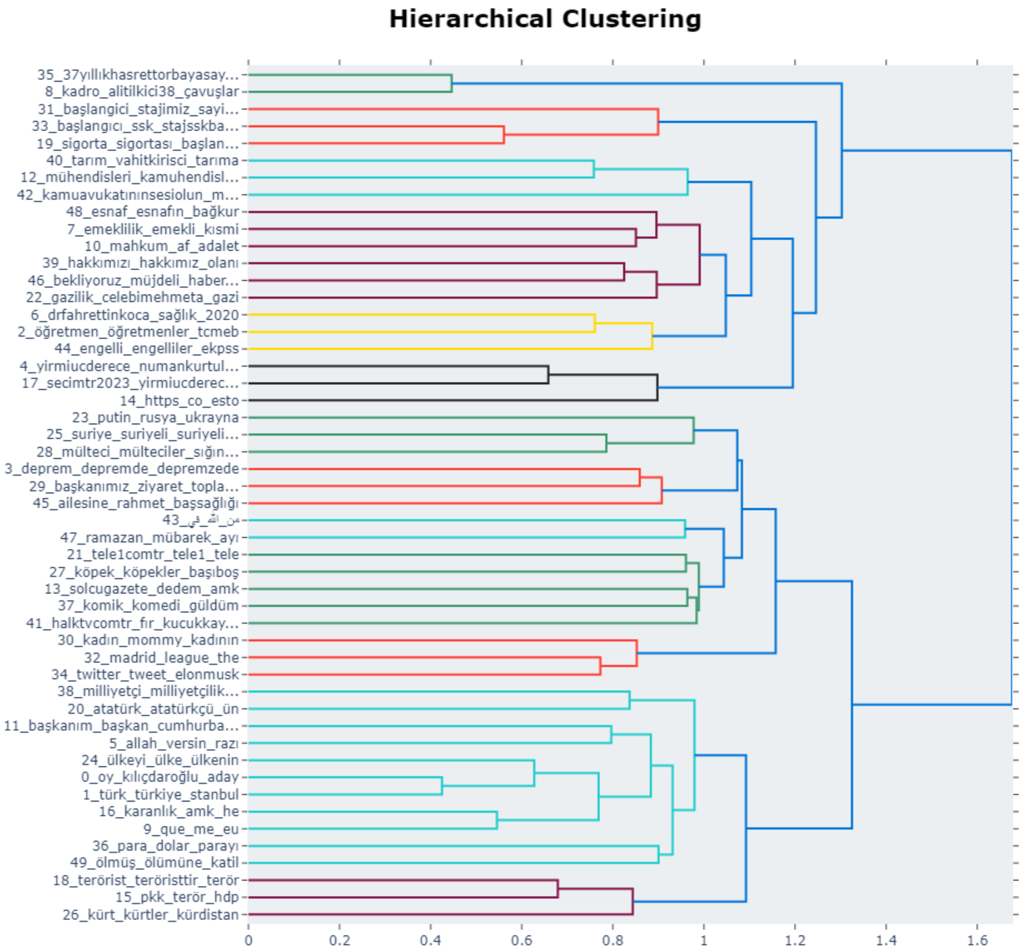

Figure 4.5: Hierarchical clustering of the top 50 topics using BERTopic functions.

covers general political discussions on Twitter.

Commencing with the bottom cluster, the green cluster illustrates topics related to foreigners, refugees, and specifically, people from Ukraine and Syria in separate subclusters. The volume of the topic related to Ukraine has stayed at a high level since the beginning of the dataset till the beginning of 2023, and the importance of the migration policy in the May 2023 elections was discussed in the previous section.

The succeeding two topic clusters, visualized as red and turquoise, demonstrate topics related to the earthquake, Arabic texts and Ramadan, respectively.

The next two clusters, colored as green and red, contain eight different partially political topics that are challenging to generalize. One topic that is considerably interesting to point out would be the topic titled as *27_köpek_köpekler_başıboş*. It refers to problems caused by free-roaming dogs on the streets of Turkey, including increased dog attacks, hygiene concerns, safety risks, and traffic accidents, worsened by uncontrolled abandonment and the expansion of these animals in public spaces.

Another topic to highlight is the topic with *30_kadın_mommy_kadının*. *kadın* can be translated to *woman* and both *kadın* and *mommy* refers to the IyiP leader Meral Akşener. Most of the young people on Twitter refer to her as a mom, underlining her warmth and sympathy.

The topic titled as *34_twitter_tweet_elonmusk* contains representative words referring to Elon Musk's acquisition of Twitter, which started around April and concluded around Oktober, the timeframe which this dataset also partially covers.

The second last cluster, portrayed as turquoise, covers various political topics from Nationalism to Atatürk, from political figures to presential candidates, from religious wishes to treason accusations, from exchange rates to other economic topics.

The last cluster, illustrated as brown, covers topics related to terrorism and the Kurdish community. One can realize the topic titled as *15_pkk_terör_hdp*, which includes in its top three representative words terrorism and Peoples' Democratic Party (HDP). HDP received around 9% of the votes in the May 2023 elections, with a significant portion of their support coming from the Kurdish community. The AKP government and other nationalists accuse HDP that it is linked to the banned Kurdistan Workers' Party (PKK), which the European Union also lists as a terrorist organization (Esen, 2022b), which we can also see from the topic representation being in the top 50 topics and its' representative tweets.

To remind it again, the extensive Turkish results and graphs, including the code, can be found on the GitHub repository of the thesis.

5 Discussion

The last chapter focused on results derived from the BERTopic topic model and analyzed them in detail. This chapter focuses on interpreting these results in a broader context and discusses this thesis's limitations and future work.

5.1 Interpreting the results in a broader context

Having the results analyzed in detail, this section focuses on interpreting these results in a more expansive concept. It aims to answer the remaining research questions about parties and their election agendas and compare the topic model results with related research. This section aims to answer the third research question.

Analyzing the 1000 topics individually, one can realize that these topics cover lots of topics, which also mirror the parties' agendas and public statements on both ruling and opposition ranks. These resulting topics cover a wide range of strategies utilized in the months leading up to the election, including those used by the opposing and ruling blocks to strengthen their position against each other. Specifically, the topics span various issues such as the rule of law, the constitution, the economy in general, both foreign and domestic policy, mottos of election campaigns, and many more.

The focus starts with the opposition rank. The Nation Alliance of six parties was formed in 2022 mainly because it focused on restoring the rule of law. They introduced the so-called ´Strengthened parliamentary system', a form of government that aims to return to the parliamentary system from the presidential system that has been in effect since the 2017 referendum. Without going into detail, the Nation Alliance seeks to limit the executive branch's powers by empowering the legislature, the Turkish parliament, to prevent the slide to autocracy in the future (Şar, 2023).

One can realize the importance of political figures in Turkish politics by looking at the top topic model results. That is why the opposition's candidacy has also been a hot topic for months, with the view that only a strong candidate could prevail against Erdoğan. The topic of opposition candidacy, being one of the top topics, supports that idea. There are almost more than 12 topics regarding political figures in the top 100 topics, which make up more than 10% of all tweets analyzed. These 12 topics refer to Kılıçdaroğlu, Erdoğan, Meral Aksener, mayors like Ekrem İmamoğlu and Mansur Yavaş, ministers like Mustafa Varank.

Introducing a ´Strengthened parliamentary system' by the Nation Alliance focuses on ensuring an independent and impartial judiciary, reforming public institutions, and systematically preventing human rights violations (Şar, 2023). One of the mottos of the opposition party CHP highlighting this issue was *'hak, hukuk, adalet'*, which translates to rights, law, and justice. There are more than four different topics in the topic model results, and they focus on the constitution, the regulations and trials, and the imprisonment sentence for İmamoğlu, the mayor of Istanbul. Most of the topics in this section refer to human rights violations and dependent and biased judiciary.

Around four topics in the topic model results relate to the imprisonment sentence for the mayor of Istanbul, İmamoğlu, and contain more than 550 thousand tweets in the dataset. The imprisonment sentence just before the elections and the justification of it are solid examples of a biased judiciary. For context, the first local election in Istanbul was canceled, and it happened again. İmamoğlu won both of the elections, with a significant difference in the second one, gaining support. In France in 2019, İmamoğlu stated in a conference that the government canceled the elections to try to win in the second round, where they used extensive state resources and polarized the people. Süleyman Soylu, who was the minister of the Interior, called Imamoğlu a fool for complaining about Turkey to the EU. On the same day, İmamoğlu called the one(s) who canceled the elections fools and said Soylu should focus on them instead (BBC News Türkçe, 2022). The court sentenced İmamoğlu to over two years imprisonment for calling Supreme Electoral Board members fools (Aktürk, 2022). Because the sentence had not yet been upheld, it was feared that if İmamoğlu were selected as the opposition candidate, the government would use his sentence to drop his candidacy, which would weaken the opposition ranks.

Another hot topic that the opposition criticized in the ruling bloc is the economic condition of the country and the ongoing hyperinflation in the months up to the elections. The unexpected earthquake made the top of the cake with massive economic repercussions (Çevik & Aksoy, 2023). Just before the elections, the official inflation rate reached a 24-year high of more than 85%, and other unofficial sources show even more than 100% (Şar, 2023).

The topic model results cover topics regarding the economy, exchange rates of the dollar and Euro to the Turkish Lira, the inflation rates, and the accused missing 418 billion dollars from the state treasury. Although these topics do not comprise the trending topics, at least five related topics can be found in the topic model results.

The exchange rate from the dollar to the Turkish Lira was 2 in 2013, 4,5 before the 2018 elections, increased to 8,5 till September 2021, and 20 before the May 2023 elections. The government has been accused of selling reserves from the treasury to keep the exchange rate low and depending on foreign money from the Gulf states and Russia (Şar, 2023). As of 2024, the exchange rate is 30, marking an increase of more than 500%

in just five years. The government also kept the central banks' interest rates low. With the hyperinflation and depleted central bank reserves, the Turkish Lira devaluation was inevitable.

The main motto of the opposition party CHP and its candidate Kılıçdaroğlu during the election campaign was *'Sana Söz yine baharlar gelecek...'*, which translates to *'Promise, the spring will come again...'*. The motto itself was also used in various campaign videos, and it originates from a song and symbolizes hope. The topic around that motto can also be found in the topic model results.

In the election month, disinformation on social media was on the rise. The campaign video of Kılıçdaroğlu based on the motto was a target of this disinformation, which was modified. An old clip of terrorists applauding and keeping rhythm was cut and added at the end of the campaign video and released on social media. In order to mobilize nationalists, the government bloc was already accusing the opposition parties of having a relationship with terrorists, and with this video becoming a trend, that thought of view skyrocketed. Even President Erdoğan showed the video during one of his rallies, which helped spread the disinformation (Oğraş, 2023).

Another topic that is deeply analyzed and discussed in this thesis is the migration policy. Since Turkey has a vast incoming migration, making Turkey the top host country for refugees in the world, the topic was also one of the trending topics of the May 2023 elections on social media. While the ruling bloc is taking a sympathetic approach to refugees, portraying them as religious brothers, the opposition bloc opposes the long-term settlement of refugees, mainly Syrians, and it promises to send the (illegal) refugees back to their countries (Esen, 2022a). The topic model results cover these points by at least four separate topics. The top topic focuses specifically on Syrians, but they do not make the top 50 topics.

The importance of the earthquake and how it affected the elections is discussed in detail in this thesis. Since the earthquake was a national disaster, it played an essential role in unifying the people in a highly polarized society. The role of unification in times of crises has also been the approach of the second biggest opposition party IyiP and its leader, Meral Akşener. She supported the idea that the earthquake was not open for a political debate and stayed silent in the first days. However, the to-be opposition candidate at that time, Kılıçdaroğlu, was present from day one in the catastrophe zone, leading a political campaign that criticized Erdoğan and his officials (Çevik & Aksoy, 2023). These two different approaches revealed profound divisions in the Nation Alliance, which continued after the selection of Kılıçdaroğlu as the joint candidate. These divisions are also some of the main reasons that push the topic of the earthquake and the opposition's candidacy, making them trending topics.

Foreign policy is another aspect to analyze. Şar (2023) remarks that foreign policy has often been seen as a part where the opposition has mainly struggled to establish a unified approach. Şar presents two reasons for that. The first one is that the parties in the Nation Alliance have different perspectives on how foreign policy should be changed or directed. Secondly, Şar states that foreign policy is typically perceived as having minimal to no impact on the voting patterns of the Turkish electorate, which gives the opposition block the motivation to focus on considerably more critical areas.

However, there is another argument that the success of the ruling bloc on foreign policy might have boosted the approval rates of Erdoğan and the People Alliance in the months up to the elections. Erdoğan achieved this by portraying himself as an influential statesman, a mediator in international conflicts, for instance between Ukraine and Russia, and a reliable defender of national interests (Çevik et al., 2023). Approval rates of approaches like this make it more complicated for the opposition block to suggest a different approach and leave them with only the choice of supporting the current foreign policy.

There are more than five different topics in the model results that cover foreign policies, the first being in the top 50 topics and about Russia and Ukraine. Other topics cover Germany and the European Union, the visa problems, NATO and the USA, and the conflict between Azerbaijan and Armenia.

Şar (2023) stated in February that for the Nation Alliance to secure a governing position, it must have first rallied an electoral majority through persuasive efforts with voters. Despite being fully aware that achieving this requires a successful electoral strategy, the alliance had, until February, focused more on developing its vision for the post-election period, rather than discussing and agreeing on a joint candidate, for instance. This highlights the main reason for the failure of the opposition bloc.

The focus of the analysis continues with the ruling ranks. It is essential to mention again that foreign policy has been one of the most influential topics that the ruling block sees as a success element, specifically after 2022, being a mediator in international conflicts, for instance, between Ukraine and Russia. The government also portrayed itself as defending national interests, increasing decisive military power in foreign countries. Military operations in Libya, Syria, Iraq, and the South Caucasus, military bases in Somalia and Qatar, and a growing domestic arms industry are solid examples (Çevik & Aksoy, 2023). Another example was the Turkish government vetoing Sweden's accession to NATO.

However, in January 2024, after US President Joe Biden sent a letter to Congress expressing the selling of F-18 fighter jets, the Turkish government approved Sweden's accession to NATO (Euronews, 2024). One can notice various topic model results covering these issues, going deep as addressing specific military operations, for instance, one done in Syria.

The relationship between Turkey and other countries in Gulf states and Russia has also been improving under this government, where Erdoğan uses foreign connections to gain financial and economic deals, keeping the 22-year AKP machine running (Çevik et al., 2023).

While projecting power internationally, the government successfully creates ´foreign enemies', presenting itself as safeguarding national interests. With the help of the control of the mainstream media, Erdoğan's command of the Turkish right and his ability to mobilize religious nationalist voters can not be denied. Looking at the top 100 topics of the topic model results, one can find more than nine topic representations covering the previously mentioned issues.

Three of these clusters cover topics around terrorism and the accusation of a relationship of parties with terrorist organizations. These are solid examples of why the opposition block had to convince people they were not affiliated with any international or terrorist organizations.

At least three clusters in the top 100 topic model results present topics related directly to President Erdoğan. While two of them show support for Erdoğan, the other prays for Erdoğan. A representative tweet for the latter topic would be *'May God keep Erdoğan in power.'*. No less than two clusters in the top 100 topic model results cover nationalism, and at least one of them is directly related to the MHP since its name includes nationalism. The representative tweets in this cluster demonstrate voting for Erdoğan and MHP in the first round of the elections, mainly aiming to increase MHP's power in the parliament.

The cluster on the 40ṗlace of the topic model result is about trust in the state, which would mean supporting the governing block. An example representative tweet here is *'We trust our state, our goal is to serve our state.'*. That is a solid example of how the ruling block successfully merged the state and the party. The merger of the state and the government, the unfair elections, and many other reasons sum up why the government can be called a competitive authoritarian regime, a concept in Comparative Politics defined by Levitsky and Way (2002).

One of the main differences between the opposition and the ruling block is their emphasis on religion. The opposition block offers a secularist approach to religion. For instance, the opposition aims to cut some costs from the Ministry of Religious Affairs, which has a yearly budget of over 2 billion dollars, exceeding many other ministries. The ruling block, on the other hand, emphasizes traditional Islam and family values, and uses religion wherever it can (Çevik et al., 2023). Kılıçdaroğlu's religious identity, Alevism, also does not contribute to convincing religious conservative voters, which diverges from the traditional Sunni Islam.

An essential argument for why some people vote for the governing block is the guarantee of continuity and stability. According to Çevik et al. (2023), the governing block was socio-politically and ideologically more coherent than the opposition and succeeded in presenting itself like that. At least one cluster in the top 50 topic model results also supports the argument.

Finally, it is essential to mention the motto of the ruling block, which was *'Türkiye Yüzyılı'*, which translates to *'Turkey's Century'*. One of the clusters in the top 5 topic representations that include at least 2,7 million tweets also covers this topic. The ruling block used this motto, which highlights the 100ẏear of the Turkish Republic, also presenting a website[1]. The website includes many (mega) projects from more than 17 categories, some covering more than 65 projects. These projects cover what the ruling block has done till the elections and their aims for the following century.

5.2 Comparing the results with relevant research

This section aims to compare and analyze the results of this thesis with other relevant research and their results and answer the last research question. While the first focus lies on research relevant to Turkish elections, the second focus is on research in other countries.

The most important paper for this thesis would be from Najafi et al. (2022), with which the dataset of this thesis was published. The research discovered that the trending topics of the dataset from July 2022 to October 2022 reflect important events from sports, political debates, or TV, and suggests that almost half of the top five daily trends are generated by coordinated attacks.

The second research from the same chair is from Najafi and Varol (2023), which developed its own language model called TurkishBERTweet, which has been trained on almost 900 million tweet data from 2009 to 2022. The chair applies the LLM on the #Secim2023 dataset for sentiment analysis of 336 million tweets. The Figure 5.1 presents daily sentiment and the sentiment difference from the mean of the tweets from July 2022 to June 2023 in one graph. One can realize some intense sentiment values in the figure. For instance, 29 October and 1 January have extreme levels of positive sentiments. 29 October is the day of the proclamation of the Turkish Republic, and the latter is New Year's Eve.

On the other hand, 6 February and 27 February mark high levels of negative sentiments. Both of the dates are about the earthquake, the first one being the day of the earthquake and the latter about the day Erdoğan visited earthquake regions and asked for forgiveness. On the same day, a trend on Twitter started against Erdogan's ask for forgiveness. This thesis's topic model results also present the earthquake topic as one of the most spoken issues.

The following two research reports are based on an EU-supported project[2] imple-

[1] https://turkiyeyuzyili.com
[2] https://cordis.europa.eu/project/id/101082050

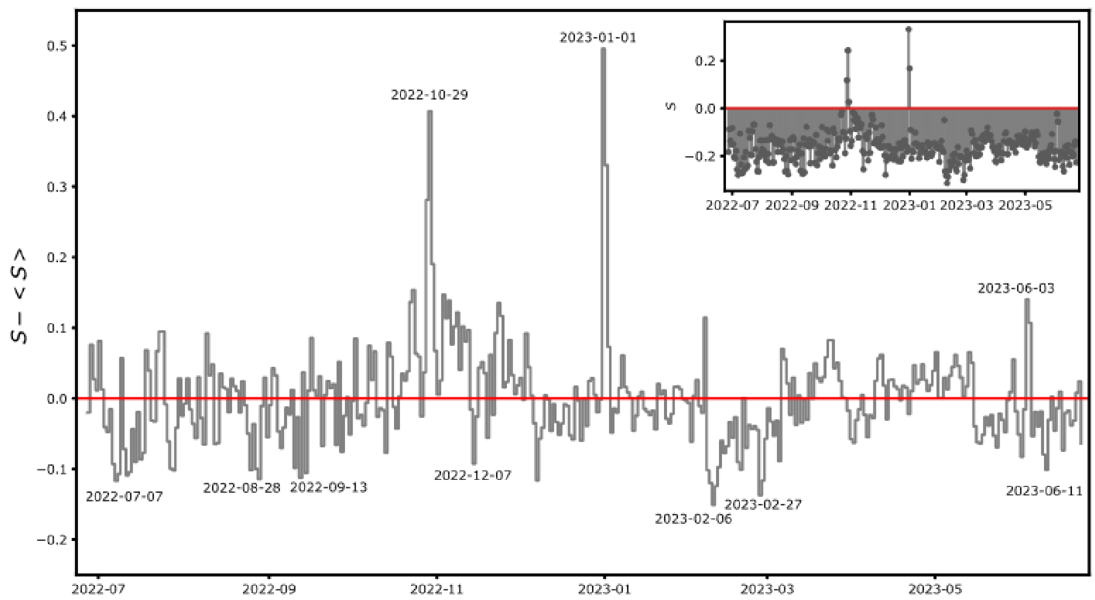

Figure 5.1: Daily sentiment, and sentiment difference from the mean. S and <S> stand for sentiment and mean sentiment of tweets. The figure is taken from the research of Najafi and Varol (2023)

mented by Koc University in Istanbul.

The first report, by Politus Analytics (2023a), presents sentiment analysis, emotional analysis, and idealogy analysis of the social media data, which includes in sum 310 million tweets.

The results propose that the primary emotion of the results was anger, highlighting the general correlation between tweeting and being angry. Naturally, after the 6 February earthquake, the dominant emotion was sadness, while happiness reduced sharply. Hope increased dramatically in the week of the candidacy announcement of Kılıçdaroğlu, the results show. Desperation was examined in the opposition ranks more significant than the ruling block ranks, losing its significance only after the candidacy announcement and continuing its trend until the elections. While confusion levels among the ruling ranks remain stable in 2023 until the elections, it is unstable among the opposition ranks with ups and downs, underlining the uncertainties with candidacy and other joint approaches.

The report also underlines the relationship of anger with topics like human rights, the economy, and justice. These mentioned points support the results of the thesis' topic model results. The emphasized emotions demonstrate why some of the discussed topics in this thesis were trending during the months up to the election.

The report also analyzes specific topics' trends between January and May. The results suggest similar results to the thesis' topic results, where the earthquake changed the trend of the discussed topics on social media in February, taking the top place. After March, the focus lies more on the upcoming election and other political discussions. Other topics cover the economy, the constitution and the regime, foreign policy, and the Kurdish community, once more akin.

The second report, by Politus Analytics (2023b), was published between the first and second rounds of the May 2023 elections and aimed to continue the first report. The analysis starts with the approval rates of Kılıçdaroğlu and Erdoğan between January and May 2023, according to Twitter, with the use of three different models. The result of the analysis suggests that the support rate of Erdoğan was, in general, higher than that of Kılıçdaroğlu, in which the gap between them peaked at the end of April.

The comparison with relevant research continues with different research papers that have used BERTopic or another topic model to analyze data from various domains, where the focus of this thesis remains as the political domain.

To start with, the first research that leveraged the power of different BERTopic models suggests that BERTopic results seem more specific than other topic models (Egger & Yu, 2022). This thesis also underlines that if the default values of parameters are not manually changed heavily so that the minimum size of clusters gets enormous, the BERTopic results appear specific. More than 7000 in the first approach and more than 1000 topic cluster results in the second approach are substantial examples. Furthermore, if one analyzes the topic clusters carefully, it is realizable that there are many similar topics. Nevertheless, these similar topics appear to differ at least in one part in many cases, for instance, one topic being about conservative nationalism and the other Kemalist nationalism. However, it is essential to highlight that analyzing the results usually takes more work when the number of topic clusters reaches more than several hundred.

The following research by Ilyas et al. (2020) used the LDA topic model to analyze the trend of topics related to Brexit on Twitter. The research had a similar research question to this thesis to determine if the discussed topic on Twitter was real-life events. The research analyzed a three-month period and found that their results highly represented the actual events. These topics cover speeches of Boris Johnson and the Queen, the deal with the EU, and events around the parliament. All of these events were trending topics during that period.

Although the analyzed time period in this thesis is four times more than this research, this thesis also found a correlation between real-life events and the discussed topics on Twitter. Some of the topics that this thesis found are the earthquake, the prison sentence of Imamoğlu, and the election results or events that happened during the elections. This correlation suggests that politicians can use social media data and their analysis to better understand public opinions and determine their policies respectfully.

The following two research focus on German elections and the topics around them on social media. The first one by Stier et al. (2018a) focuses on the German federal election campaign in 2013 and trains a Bayesian language model to identify topics from politicians' posts and their comments. The research suggests that the messages from politicians and their audiences prioritize different topics. As future work, this thesis could also separate tweets from the politicians and voters and analyze the corresponding results accordingly to find the difference between them in Turkish politics. Stier et al. (2018a) discovers that the top six topics covered on Twitter were political debates, polity, law and order, coalition formation, campaigning, and the labor market. All of the mentioned topics in the 2013 German elections were similar to these thesis findings of the 2023 Turkish elections. However, unlike the 2023 Turkish Twitter political discourse, in German political discourse, the following topics are very short in volume: economy, currency and Euro, migration and integration, and foreign policy.

The following research by Gritto (2022) is based on a thesis where BERTopic and other topic models were applied to Twitter data written by German politicians in the months up to the 2017 German federal elections. Since the focus lies on politicians, the number of tweets analyzed is considerably lower than this thesis. However, a significantly similar analysis strategy to this thesis was used because the analyzed tweets are in German. When the top 20 topics of the German politicians are analyzed, some similarities and differences are noticed with the Turkish political discourse. Similarities lie on topics like thankfulness, police, the chancellorship duel, election campaigns, freedom of the press, foreign affairs, armed forces, and pensions. These mentioned topics can be found in various clusters with minimal contrasts in the topic model results in this thesis. On the other hand, topics like climate change and digitization are in the top four topics discussed by German politicians, which are not even noticeable in the Turkish political discourse. One solid argument for this comparison would be the difference in support for the Green parties in Germany and Turkey, where the Green party is not even in the Turkish parliament.

The next research by Stier et al. (2018b) focuses on another developing country, Brazil, and analyzes data from political websites using other topic modeling approaches than BERTopic. The most exciting aspect of analyzing the results of this research would be to find the differences between the results from German political discourse and the results from Brazil and maybe also Turkey. Unlike the two previous German elections, this research suggests that the economy is one of the top discussed topics, similar to this thesis' results but differing from the German election results. It is essential to note that the data collection method differs in these research papers. Another interesting aspect would be the topics covering nationalism and religion found in both Turkish political discourse and the Brazilian political websites but not in tweets from German politicians.

The last aspect of comparison is research based on the 2016 US elections. The first research is a comparative politics paper by Theocharis et al. (2020) focusing on incivilities on Twitter. However, the research also does topic model analysis leveraging LDA on tweets by members of the US Congress, similar to the thesis of Gritto (2022). Although there are no topic representations, one can understand the topics from the top representative words. The topics cover important figures like Donald Trump, Hillary Clinton, Barack Obama, and Martin Luther King, as well as foreign policy clusters mentioning Syria, Israel, Afghanistan, and Iran. In addition to that, illegal immigration, elections and voting, gun rights and violence, the military, tax, and jobs make the top topic results. Most of the mentioned topics align with the results of this thesis, diverging in a few topics like gun rights, drugs, and climate.

Both Fang et al. (2019) and Fang (2019) focus on the 2016 US elections and analyzing Twitter data. The big difference is that the research divides the data into two significant clusters, one supporting Clinton and the other against Trump. The research suggests that the clusters include opposing topics in each other and highlights that both clusters cover controversial topics that are popular among both communities. By dividing the data into two clusters, the research covers more specific topics like email leaking (Wikileaks), Mexican immigrants, and the border wall, among the top topics, which can also be realized compared with the results of this thesis. On the other hand, unlike the previous research, topics like gun rights, drugs, and climate can not be found in this research.

The following research from Yaqub et al. (2017) has the same research focus as this thesis, trying to find a correlation between real-life events and popular topic discussions on Twitter. The research focuses on the most popular terms and analyzes their daily occurrence and peaks. The results show that popular terms like FBI, Email, Wikileaks, Obama, and Protest were indicated election-related events, discussions, and news. These results align with the results of this thesis, highlighting the use of social media data for public opinion analysis. Another significant result of this research was the content creation amongst Twitter users. The research highlights that Twitter was primarily used for rebroadcasting already present opinions by retweeting with little communication between users, which is also aligning with the dataset of this thesis and other research, for instance, from Pfeffer et al. (2023).

5.3 Limitations

This section focuses on mentioning the limitations of the thesis.

First, the collection of #Secim2023 dataset created a problem due to the lack of official academic API and usage of other routes, which led to a partial dataset collection. Since the collection started after October 2023, a part of the dataset was also deleted. These factors led to the collection of 150 million tweets out of 250. Another aspect to consider is the monthly distribution of the #Secim2023 dataset, where almost one-third of the collected data are from March 2023, which could bias the results into making the top results topics from the trends in March. This thesis analyzes the topic clusters percentage-wise and monthly to partially overcome this aspect.

The root of the thesis dataset, Twitter data, also has its own limitations. The research results by Elmas et al. (2021) suggest that at least 20% of the top ten global trends and at least 47% of the top five daily trends in Turkey are created from automated attacks. Another research by Najafi et al. (2022) uses Botometerlike, presented by Yang et al. (2022), which can be used to evaluate Twitter accounts for finding bots. The results suggest that popular accounts usually have more than 50% bot followers, which can work for or against political campaigns.

In the results of the topic model, it can be seen that, in some cases, all of the representative tweets of a topic are almost similar, creating a trend and interaction from that, making the topic appear in the results of this thesis. Further research suggests that almost 80% of all tweets refer to others in the form of retweets or quotes, which also underlines this issue (Pfeffer et al., 2023). The topics created from automated attacks vary from support for an election campaign to backing a demand from the government.

Additional research finds out that 90% of all tweets are sent by around 4% of total users on Twitter, 75% of all tweets by around 1,7% (Pfeffer et al., 2023). In other words, this implies that this thesis analyzed a specific subset of Twitter users, which may slightly restrict the generalizability of the thesis findings.

The limitations of the Twitter data continue with the length of individual tweets. Since around February 2023, each tweet had a character limit of 280. It was then increased to 4000 in February 2023 and 10,000 in April 2023, at the end of the #Secim2023 dataset. Unlike other text data, tweets usually tend to be short, which could make it harder for BERTopic to cluster the tweets into a topic. The character increase after 2023 could also affect the analysis and the resulting trends.

The second aspect of the limitation would be from the topic model, BERTopic, although it is one of the state-of-the-art neural topic models. Since the dataset is in Turkish, a multilingual sentence embedding model embeds the documents in this thesis. However, the multilingual models have lower average performance than English models[3].

Further research also suggests that the embedding approach results in many topics,

[3]https://www.sbert.net/docs/pretrained_models.html

which requires effort demanding examination of each topic (Egger & Yu, 2022). The thesis results also align with the findings since in one approach, 1000, and in another, more than 7000 topic clusters are found. Some approaches to overcoming this issue will be highlighted in the next chapter. However, it is essential to mention that establishing an optimal number of topics is problematic because most topics overlap and cover a mixture of two to three different aspects. Although that could be troubling for some analysis, in this thesis, having different aspects of one topic can explain the stance towards the topics, clarifying the relationship to the government or the opposition.

While labor-intensive analysis of each topic is necessary, objective evaluation is also missing in this approach. Also in this thesis, the underlying meanings are subject to human interpretation. During the analysis of the results, the necessity in many cases to look at more than five representative tweets was not available. That was also the case because, in numerous instances, most representative words are redundant, making it almost impossible to interpret the topic cluster. The topic *'Criticism of Political Figures'* has the following representative words *'siyaset, siyasetin, siyasetiniz, siyasetçi, siyaseti, siyasete, siyasette'*, making it a perfect example.

5.4 Future Work

Citation test (Lamport, 1994).

6 Conclusion

6.1 Section

Citation test (Lamport, 1994).

Acronyms must be added in `main.tex` and are referenced using macros. The first occurrence is automatically replaced with the long version of the acronym, while all subsequent usages use the abbreviation.

E.g. `\ac{TUM}, \ac{TUM}` ⇒ Technical University of Munich (TUM), TUM

For more details, see the documentation of the `acronym` package[1].

6.1.1 Subsection

See Table 6.1, Figure 6.1, Figure 6.2, Figure 6.3.

Table 6.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

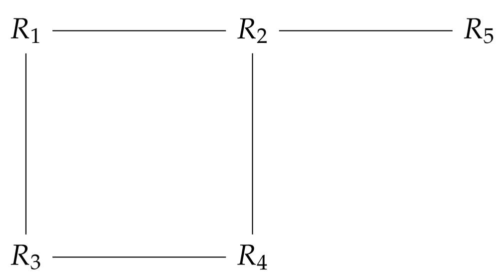

Figure 6.1: An example for a simple drawing.

[1] https://ctan.org/pkg/acronym

Topic Label	Representative Words
Elections & Candidates	oy, kılıçdaroğlu, aday, seçim, erdoğan, tayyip, seçimi, istifa
Erdogan & Political Developments	türk, türkiye, stanbul, yüzyılı, turkey, türkiyeyüzyılı, ankara, cumhuriyeti
Political Agenda & Demands	yirmiucderece, numankurtulmus, aysedogan1955, cenginyurt52, herkesicinchp, meral_aksener, secimtr2023, hassa61
Religious Wishes & Political Figures	allah, versin, razı, eylesin, müslüman, etsin, din, rabbim,
Teachers & Demands	öğretmen, öğretmenler, tcmeb, ataması, prof_mahmutozer, 100, bin, kpss
Earthquake & Demands	deprem, depremde, depremzede, depremin, yapıkayıt, yumuşak, depremden, müstakil
Nation & Country	ülkeyi, ülke, ülkenin, ülkeye, ülkede, millet, milletin, milleti
Leading Figures	başkanım, başkan, cumhurbaşkanım, mansuryavas06, ekrem_imamoglu, başbakan, cumhurbaşkanı, sayın
Fair Trial & Amnesty Demands	mahkum, af, adalet, genelaf, 77, adil, adli, mahkumlar
Mixed Emotions	que, me, eu, não, no, dedem, aq, amk
Civil Servants & Demands	sırtlayan, hafızası, kurumların, yükünü, kamunun, 3600ekgösterge, devletine, umudumuz
Retirement System	emeklilik, emekli, kısmi, kademeli, prim, 5000, yaş, zorunlu
Reserves Demands	degildi, tercihimiz, dileğimiz, milletvekilim, etmektir, yedek, talebimiz, mehmetfatihser5
Medical Secretaries & Demands	drfahrettinkoca, sağlık, 2020, tıbbi, sağlıkçılar, sekreterlik, sağlıkçı, yönetimi,
Election News & Comments	secimtr2023, yirmiucderece, cumhuriyetgzt, vekilince, https, co, ozan_blk07, furkancerkes

Table 6.2: Result of the topic modeling with reduced topic distribution. The top eight representative words from MMR model for each of the top 15 topics with their respective topic label translated into English.

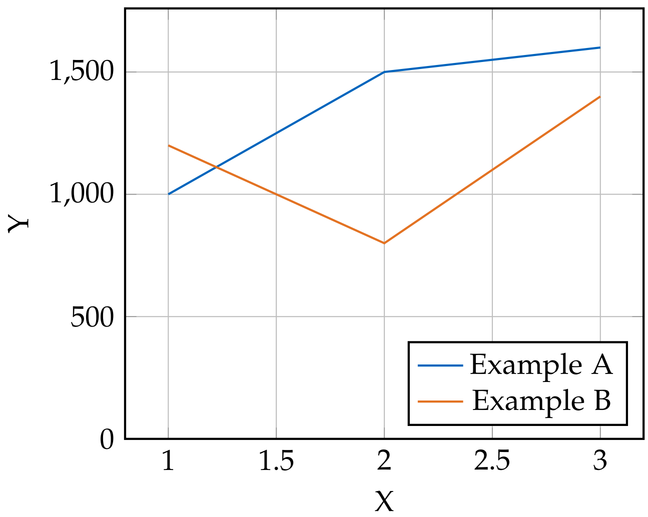

Figure 6.2: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 6.3: An example for a source code listing.

Abbreviations

TUM Technical University of Munich

CHP Republican People's Party

AKP Justice and Development Party

MHP Nationalist Movement Party

HDP Peoples' Democratic Party

DEVA Democracy and Progress Party

IyiP Good Party

ZP Victory Party

SAADET Felicity Party

DP Democrat Party

DEVA Democracy and Progress Party

GP Future Party

NLP Natural Language Processing

PKK Kurdistan Workers' Party

List of Figures

List of Tables

Bibliography

Aktürk, A. C. (2022). İBB Başkanı İmamoğlu'na verilen hapis cezasının gerekçesi açıklandı. Retrieved March 1, 2024, from https://www.aa.com.tr/tr/gundem/ibb-baskani-imamogluna-verilen-hapis-cezasinin-gerekcesi-aciklandi/2774662

Alshaabi, T., Dewhurst, D. R., Minot, J. R., Arnold, M. V., Adams, J. L., Danforth, C. M., & Dodds, P. S. (2021). The growing amplification of social media: Measuring temporal and social contagion dynamics for over 150 languages on Twitter for 2009-2020. *EPJ data science*, *10*(1), 15. https://doi.org/10.1140/epjds/s13688-021-00271-0

Anwar, A., Ilyas, H., Yaqub, U., & Zaman, S. (2021). Analyzing QAnon on twitter in context of US elections 2020: Analysis of user messages and profiles using VADER and BERT topic modeling. *DG.O2021: The 22nd Annual International Conference on Digital Government Research*, 82–88. https://doi.org/10.1145/3463677.3463718

Ardan, M. (2023). 1994 Financial Crisis in Turkey. In B. Açıkgöz (Ed.), *Black Swan: Economic Crises, Volume II* (pp. 95–126). Springer Nature. https://doi.org/10.1007/978-981-99-2318-2_7

Atila, S. (2022). 3 kasım 2002'den bugüne akp ve erdoğan'ın 20 yıllık seçim tarihi. Retrieved February 3, 2024, from https://medyascope.tv/2022/11/03/3-kasim-2002den-bugune-akp-ve-erdoganin-20-yillik-secim-tarihi/

BBC News Türkçe. (2022). İmamoğlu 2 yıl 7 ay 15 gün hapis cezasına çarptırılması sonrası konuştu: 2023 ideallerimiz var. Retrieved March 1, 2024, from https://www.bbc.com/turkce/articles/c4n2yzp3d7vo#:~:text=%C4%B0stanbul%20B%C3%BCy%C3%BCk%C5%9Fehir%20Belediye%20(%C4%B0BB)%20Ba%C5%9Fkan%C4%B1,15%20g%C3%BCn%20hapis%20cezas%C4%B1%20verildi.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.5555/944919.944937

Calma, J. (2023). Twitter just closed the book on academic research. Retrieved February 7, 2024, from https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research

Caramani, D. (2020). *Comparative Politics*. Oxford University Press.

Çevik, S. (2018). The Future of Opposition in Turkey. Overcoming Identity Politics Is the Key for Success. Retrieved February 25, 2024, from https://www.swp-berlin.org/publikation/the-future-of-opposition-in-turkey

Çevik, S., & Aksoy, H. A. (2023). Political and economic implications of the Turkish earthquakes. https://doi.org/10.18449/2023C19

Çevik, S., Aksoy, H. A., & Aydın, Y. (2023). Turkey's New Cabinet: A Wind of Change in Turkish Politics? https://doi.org/10.18449/2023C42

Contreras, K., Verbel, G., Sanchez, J., & Sanchez-Galan, J. E. (2022). Using topic modelling for analyzing panamanian parliamentary proceedings with neural and statistical methods. *2022 IEEE 40th Central America and Panama Convention (CONCAPAN)*, 1–6. https://doi.org/10.1109/CONCAPAN48024.2022.9997766

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.886498

Elmas, T., Overdorf, R., Özkalay, A. F., & Aberer, K. (2021). Ephemeral Astroturfing Attacks: The Case of Fake Twitter Trends [arXiv:1910.07783 [cs]]. https://doi.org/10.48550/arXiv.1910.07783

Erdem, E., & Bişgin, E. (2023). Turkey - Changes to Minimum Retirement Age Requirement. Retrieved February 27, 2024, from https://kpmg.com/xx/en/home/insights/2023/06/flash-alert-2023-121.html

Esen, B. (2022a). Post-2023 election scenarios in turkey. https://doi.org/10.18449/2022C55

Esen, B. (2022b). The Opposition Alliance in Turkey: A Viable Alternative to Erdoğan? https://doi.org/10.18449/2022C52v02

Euronews. (2024). Biden Kongre'ye F-16 mektubu gönderdi, Türkiye İsveç'in NATO üyelik sürecini tamamladı. Retrieved March 4, 2024, from https://tr.euronews.com/2024/01/25/biden-kongreye-f-16-mektubu-gonderdi-turkiye-isvecin-nato-uyelik-surecini-tamamladi

Fang, A. (2019). *Analysing political events on twitter: Topic modelling and user community classification* (Doctoral dissertation). University of Glasgow. https://doi.org/10.5525/gla.thesis.41135

Fang, A., Habel, P., Ounis, I., & MacDonald, C. (2019). Votes on twitter: Assessing candidate preferences and topics of discussion during the 2016 u.s. presidential election [Publisher: SAGE Publications]. *SAGE Open*, *9*(1), 2158244018791653. https://doi.org/10.1177/2158244018791653

Freedom House. (2023). *Turkey* (tech. rep.). Retrieved February 20, 2024, from https://freedomhouse.org/country/turkey/freedom-world/2023

Gritto, A. (2022). *Application of neural topic models to twitter data from German politicians* (bat). Ludwig-Maximilians-Universität München. https://doi.org/10.5282/ubm/epub.92617

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. https://doi.org/10.48550/arxiv.2203.05794

Gültekin, L. (2023). Kılıçdaroğlu'nun adaylığını niçin riskli buluyorum? - Diken. Retrieved February 28, 2024, from https://www.diken.com.tr/kilicdaroglunun-adayligini-nicin-riskli-buluyorum/

Gündoğan, B., Yıldızalp, M., & Tosun, M. (2023). CHP'nin yeni genel başkanı Özgür Özel oldu. Retrieved February 28, 2024, from https://www.aa.com.tr/tr/politika/chpnin-yeni-genel-baskani-ozgur-ozel-oldu-/3044068

Haber, T. (2023). Kılıçdaroğlu: Cumhuriyet'in 100. yılında 100 bin öğretmen ataması yapacağız. Retrieved February 28, 2024, from https://www.trthaber.com/haber/gundem/kilicdaroglu-cumhuriyetin-100-yilinda-100-bin-ogretmen-atamasi-yapacagiz-765768.html

Ilyas, S. H. W., Soomro, Z. T., Anwar, A., Shahzad, H., & Yaqub, U. (2020). Analyzing brexit's impact using sentiment analysis and topic modeling on twitter discussion. *The 21st Annual International Conference on Digital Government Research*, 1–6. https://doi.org/10.1145/3396956.3396973

Kaiser, J., Rauchfleisch, A., & Bourassa, N. (2020). Connecting the (far-)right dots: A topic modeling and hyperlink analysis of (far-)right media coverage during the US elections 2016. *Digital Journalism*, *8*(3), 422–441. https://doi.org/10.1080/21670811.2019.1682629

Klosowski, T. (2022). Why You should Delete (All) your tweets. Retrieved February 20, 2024, from https://www.nytimes.com/wirecutter/blog/why-you-should-delete-your-tweets/

Lamport, L. (1994). *Latex : A documentation preparation system user's guide and reference manual*. Addison-Wesley Professional.

Levitsky, S., & Way, L. A. (2002). Elections without democracy: The rise of competitive authoritarianism. *Journal of Democracy*, *13*(2), 51–65. https://doi.org/10.1353/jod.2002.0026

McInnes, L., Healy, J., & Melville, J. (2020, September 17). UMAP: Uniform manifold approximation and projection for dimension reduction. https://doi.org/10.48550/arXiv.1802.03426

Mejova, Y., Weber, I., & Macy, M. W. (Eds.). (2015). *Twitter: A digital socioscope*. Cambridge University Press. https://doi.org/10.1017/CBO9781316182635

Michaelson, R., & Narlı, D. B. (2023). 'He works hard': Voters in Turkey's quake zone backing Erdoğan in runoff. Retrieved February 27, 2024, from https://www.theguardian.com/world/2023/may/26/turkey-quake-zone-voters-backing-erdogan-in-runoff

Najafi, A., Mugurtay, N., Demirci, E., Demirkiran, S., Karadeniz, H. A., & Varol, O. (2022). #Secim2023: First Public Dataset for Studying Turkish General Election. Retrieved October 28, 2023, from http://arxiv.org/abs/2211.13121

Najafi, A., & Varol, O. (2023). TurkishBERTweet: Fast and Reliable Large Language Model for Social Media Analysis. https://doi.org/10.48550/arXiv.2311.18063

Ogan, C., & Varol, O. (2017). What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement: Twitter use during gezi park. *Information, Communication & Society*, *20*(8), 1220–1238. https://doi.org/10.1080/1369118X.2016.1229006

Oğraş, M. (2023). Millet İttifakı kampanya filminde Murat Karayılan'ın yer aldığı iddiası - Teyit. Retrieved March 3, 2024, from https://teyit.org/demec-kontrolu/millet-ittifaki-kampanya-filminde-murat-karayilanin-yer-aldigi-iddiasi

Pfeffer, J., Matter, D., Jaidka, K., Varol, O., Mashhadi, A., Lasser, J., Assenmacher, D., Wu, S., Yang, D., Brantner, C., Romero, D. M., Otterbacher, J., Schwemmer, C., Joseph, K., Garcia, D., & Morstatter, F. (2023, April 11). Just another day on twitter: A complete 24 hours of twitter data. http://arxiv.org/abs/2301.11429

Politus Analytics. (2023a). Politus Türkiye Panoraması 1: Seçimlere Giderken Hangi Seçmen Ne Hissediyor ve Neyi Konuşuyor. https://politusanalytics.com/portfolio/politus-turkiye-panoramasi/

Politus Analytics. (2023b). Politus Türkiye Panoraması 2: Birinci Tur Sonrası Seçmen Eğilimleri. https://politusanalytics.com/portfolio/politus-turkiye-panoramasi-ii/

Rabasa, A., & Larrabee, F. S. (2008). *The Rise of Political Islam in Turkey* (1st ed.). RAND Corporation. https://www.jstor.org/stable/10.7249/mg726osd

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Retrieved November 7, 2023, from http://arxiv.org/abs/1908.10084

Reimers, N., & Gurevych, I. (2020, October 5). Making monolingual sentence embeddings multilingual using knowledge distillation. https://doi.org/10.48550/arXiv.2004.09813

Saç, E., & Çoban, T. (2023). Seçim sonuçlarından geriye bakmak: Anketler neden, nasıl yanıldı? Retrieved July 24, 2023, from https://teyit.org/dosya/secim-sonuclarindan-geriye-bakmak-anketler-neden-nasil-yanildi

Şar, E. (2023). The Post-Erdoğan Vision of Turkish Opposition: Opportunities and Limitations. https://doi.org/10.18449/2023C07

Schweter, S. (2020). BERTurk - BERT models for Turkish. https://doi.org/10.5281/zenodo.3770924

Scott, M. (2023). How Turkey's Erdoğan uses social media to cling onto power. Retrieved February 3, 2024, from https://www.politico.eu/article/recep-tayyip-erdogan-elon-musk-twitter-turkey-elections-social-media-power/

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018a). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter [Publisher: Routledge]. *Political Communication*, *35*(1), 50–74. https://doi.org/10.1080/10584609.2017.1334728

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018b). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter [Publisher: Routledge]. *Political Communication*, *35*(1), 50–74. https://doi.org/10.1080/10584609.2017.1334728

Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The dynamics of political incivility on twitter. *SAGE Open*, *10*(2), 2158244020919447. https://doi.org/10.1177/2158244020919447

Ulaş. (2023). CHP'li Killik'in, parti önünde 'Kılıçdaroğlu aday olmasın' pankartı açan gence tepkisi tartışma yarattı. Retrieved February 28, 2024, from https://onedio.com/haber/chp-li-killik-in-parti-onunde-kilicdaroglu-aday-olmasin-pankarti-acan-gence-tepkisi-tartisma-yaratti-1126170

Uysal, N., & Schroeder, J. (2019). Turkey's twitter public diplomacy: Towards a "new" cult of personality. *Public Relations Review*, *45*(5), 101837. https://doi.org/10.1016/j.pubrev.2019.101837

Yang, K.-C., Ferrara, E., & Menczer, F. (2022). Botometer 101: Social bot practicum for computational social scientists. *Journal of Computational Social Science*, *5*(2), 1511–1528. https://doi.org/10.1007/s42001-022-00177-5

Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*, *34*(4), 613–626. https://doi.org/10.1016/j.giq.2017.11.001

Yerlikaya, T., & Toker, S. (2020). Social media and fake news in the post-truth era: The manipulation of politics in the election process. *Insight Turkey*, 177–196. https://doi.org/10.25253/99.2020222.11

Yilmaz, I., & Bashirov, G. (2018). The AKP after 15 years: Emergence of Erdoganism in Turkey. *Third World Quarterly*, *39*(9), 1812–1830. https://doi.org/10.1080/01436597.2018.1447371

Zaharna, R. S., & Uysal, N. (2016). Going for the jugular in public diplomacy: How adversarial publics using social media are challenging state legitimacy. *Public Relations Review*, *42*(1), 109–119. https://doi.org/10.1016/j.pubrev.2015.07.006

Zhao, H., Phung, D., Huynh, V., Jin, Y., Du, L., & Buntine, W. (2021, February 28). Topic modelling meets deep neural networks: A survey. https://doi.org/10.48550/arXiv.2103.00498