Text Mining and Social Media Mining

A Deep Dive into Erdoğan's Speeches through Text Mining Analysis

Nurdan Beşli, 457945

Maciej Lorens, 419763

Huseyin Polat, 437969

February, 2024

Table of Content

1. Introduction

The power of words cannot be underestimated in the domain of political discourse. Every world leader uses their own style of rhetoric to communicate their ideologies and shape the public's opinion in their favor, with varying levels of success. In the current landscape of global leadership, few figures are as compelling to analyze as Recep Tayyip Erdoğan, the charismatic and controversial President of Türkiye. His speeches have not only secured him a long standing political career, but have also put him into the international spotlight. Utilizing advanced text mining techniques, we delved into Erdoğan's political speeches in order to discover hidden patterns and sentiments within his public addresses.

In our research, we specifically focused on confirming several unexamined assumptions about the chosen topic:

1. **Nationalistic Sentiments:** Erdoğan's speeches frequently emphasize nationalistic sentiments, underscoring Türkiye's historical importance and global standing. This rhetoric correlates strongly with his slogan 'One Nation, One Flag, One Homeland, One State', reinforcing national identity and unity among the Turkish populace.
2. **Rhetorical Shifts During Elections and Crises:** Erdoğan's rhetoric notably shifts during election periods and national crises like the 2016 coup attempt or the 2023 earthquake. These times there's a distinct change in his speech patterns compared to other periods.
3. **Sentiment Dynamics in Political Speeches:** The sentiment in Erdoğan's speeches shifts with the political context. Positive sentiment prevails during presidential or AK Party's campaigns, while negative sentiment emerges in responses to national crises like the 2016 coup and the 2023 earthquake.

2. Data

2.1. Data description

The data for this analysis originates from the public speeches delivered by Recep Tayyip Erdoğan, accessible on the official website of the President of Türkiye (https://www.tccb.gov.tr/receptayyipErdoğan/konusmalar/). All speeches available from August 28th, 2014, to December 16th, 2023, were extracted using web scraping with

Beautiful Soup. A subset of 49 speeches was carefully selected, each speech aligning with significant events in Turkish politics such as the presidential elections, the 2016 coup attempt, and the 2023 earthquake. Consequently, speeches made during election campaigns were selected both before and after the election date, while speeches delivered in the aftermath of critical events like the coup attempt and earthquake were also included in the analysis. These specific instances represent crucial tests of a politician's capacity to assert authority over the public, occurring either around the time of an election or during a crisis.

The table below provides an overview of the selected significant dates, associated events, and the number of speeches parsed before and after each event.

Event	Year	Date	# of speeches		
			before event	after event	total
President Election	2014	10 August	0	1	1
Local Elections 1st Phase	2015	7 June	4	1	5
Local Elections 2nd Phase	2015	1 Nowemb	2	1	3
Military Coup Attempt	2016	15 July	0	5	5
Referandum	2017	16 April	2	1	3
Parliamentary and Presidential Elections	2018	24 June	4	2	6
Local Elections 1st Phase	2019	31 March	4	1	5
Local Elections 2nd Phase - for only Istanbul	2019	23 June	4	2	6
Eartquake	2023	6 February	0	5	5
Parliamentary and Presidential Elections 1st Phase	2023	14 May	4	0	4
President Election 2nd Phase	2023	28 May	4	2	6

2.2. Data cleaning

All speeches, originally in Turkish, underwent translation using Google Translator from the deep_translate package. Subsequently, translations were manually verified by the authors, some of whom are native Turkish speakers. This meticulous process ensured that each translation faithfully conveyed the intended meaning. Furthermore, the text underwent cleaning procedures, including converting all letters to lowercase, eliminating English stopwords, punctuation, whitespace characters, and numbers. As a final step, we applied lemmatization using the WordNet lemmatizer.

3. Modeling

Each of the assumptions was checked with a mix of appropriate text-mining techniques. We settled on text clustering, sentiment analysis followed by topic modeling and analysis of text vectors obtained with word2vec.

3.1. Text clustering

Clustering with K-means requires representing each speech as a vector. This can be achieved by computing the TF-IDF of the tokens in each speech and constructing a vector of the values. These vectors can be used to cluster similar speeches together. To determine the appropriate number of clusters, SSE and Silhouette Score plots were visually inspected. In both cases, the best k number of clusters seems to be 7. In the SSE plot, the elbow forms around k=7 and the silhouette score is the highest for this number as well.

3.2. Sentiment analysis and topic modeling

When applying sentiment analysis on the whole speech text, our concern was that a large majority of them will overall present a positive sentiment. That is, because a politician will most likely avoid upsetting the public, instead trying to use uplifting language even in the most dire of situations. Instead, we decided to obtain the sentiment of every paragraph longer than 75 tokens (the 5th quantile of paragraph lengths) and divide them into positive and negative ones. This way, we got 529 paragraphs with a negative sentiment and 1377 with a positive one. Even though the disproportion is quite clear, both subsets are large enough to perform analysis on. The lexicon based tool Valence Aware Dictionary and sEntiment Reasoner (VADER) was used for calculating the compound score of sentiment. Then, a threshold of above 0.05 and under -0.05 was used for classifying a paragraph as positive or negative.

The model chosen for determining the topics is Latent Dirichlet Allocation (LDA). Through cross validation, models with the highest coherence score were picked for analysis of positive and negative paragraphs respectively. Coherence score is a measure used for determining the interpretability of LDA topics to humans.

3.3. Word2Vec

Skip-gram is a popular algorithm used in natural language processing and machine learning for generating word embeddings, particularly within the Word2Vec framework. Word embeddings are numerical representations of words that capture semantic relationships between them. In the skip-gram model, the goal is to predict the context words (words surrounding a target word) given a target word. The obtained embeddings can be used to find words with similar meanings or used in similar contexts.

Implementation of word2vec in gensim requires operating on sentences, so the speeches were split into sentences and the same cleaning procedure was applied on them. The chosen vector size and window size were 100 and 5 respectively.
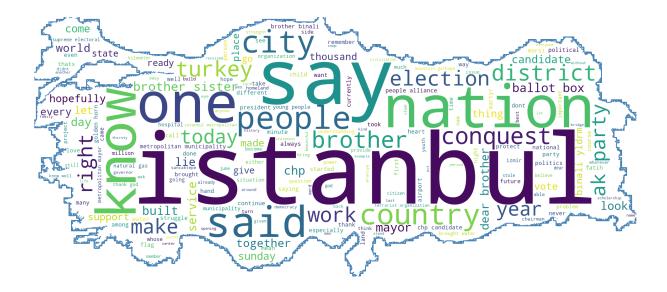
4. Evaluation

4.1. Text clustering

Cluster 0

Cluster 0 conversations consistently revolve around the topic of Istanbul, a recurring theme that holds particular significance. Out of the six speeches analyzed in this cluster, five were delivered during the pivotal 2019 Türkiye local elections, while the remaining speech was presented at the celebrations marking the 562nd Anniversary of the Conquest of Istanbul.

The 2019 Türkiye local elections, especially the Istanbul mayoral race, were crucial due to their impact on the country's political landscape. Initially won by Ekrem İmamoğlu of the opposition CHP, the election marked a significant loss for President Erdoğan's ruling AKP, breaking their long-standing dominance in Istanbul. The election in Istanbul was reorganized amid allegations of irregularities, and İmamoğlu won again, highlighting the AKP's diminishing hold on the city and signaling potential shifts in Turkish politics. Istanbul was crucial for Erdoğan because it not only served as Türkiye's economic and cultural center but also symbolized his party's long-standing political dominance and influence in the country.

Cluster 1

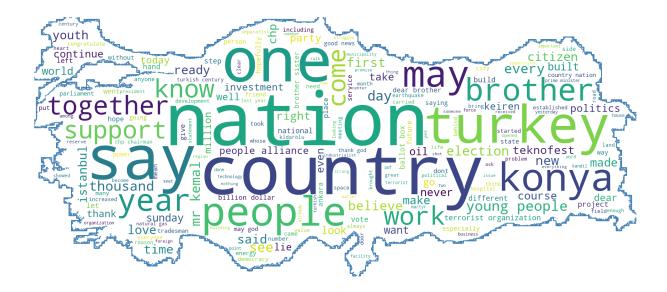On the night of July 15, 2016, Türkiye experienced a pivotal moment in its history as a coup attempt unfolded. Ultimately, President Erdoğan addressed the nation through a FaceTime call, urgently calling upon the citizens to take to the streets and resist the coup plotters, and the government was able to regain control. In the aftermath, there were widespread arrests, purges, and changes in the military and government institutions as Türkiye grappled with the implications of the failed coup. This significant event had profound implications for Türkiye's political landscape, leading to increased tensions, changes in policies, and an enduring national conversation about the role of the military in Turkish politics.

Hence, when analyzing the speeches delivered in the aftermath of the July 15, 2016 coup attempt, a distinctive pattern emerges, with all of them consistently categorized within cluster 1, setting them apart from other speeches. Notably, a thematic shift becomes evident in the prevalent use of specific terms related to the coup attempt, including words like "coup," "July," "coup attempt," "armed," "terrorist," "force," and "tank." Moreover, a remarkable development is the prominent emphasis on the word "nation" in President Erdoğan's speeches following the coup attempt, signifying a significant shift in his discourse and highlighting the profound impact of this event on the nation's collective identity and political discourse.

Cluster 2

Cluster 2 encompasses speeches delivered in the lead-up to the 2023 Turkish Presidential Election, showcasing a clear thematic focus on President Erdoğan's mission and campaign messaging. In these speeches, President Erdoğan consistently emphasized his slogan, "One Nation, One Flag, One Homeland, One State," underlining the core of his electoral strategy. The 2023 Turkish presidential election, held in two rounds on May 14 and May 28, featured three candidates initially. As no candidate secured an absolute majority of valid votes in the first round, the second round was a head-to-head contest between Recep Tayyip Erdoğan and Kemal Kılıçdaroğlu.

Distinctive keywords in Cluster 2, such as "Mr. Kemal" and "alliance," signify references to the strongest rival candidate and the political alliances competing against Erdoğan's party. Notably, Erdoğan adopted a more inclusive and conciliatory tone in his election campaign, focusing on overarching projects, party mission, and messages about his competitors.

Cluster 3

The 2017 Turkish constitutional referendum was a pivotal event in Türkiye's political history. Held on April 16, 2017, it sought to approve a series of constitutional changes that would transition Türkiye from a parliamentary system to a presidential system, significantly expanding the powers of the presidency. The referendum resulted in a narrow victory for the "Yes" campaign. The outcome signaled a significant shift in Türkiye's political landscape, consolidating President Erdoğan's authority and bringing about changes in the country's governance structure.

In evaluating the speeches associated with Cluster 3, a distinctive pattern emerges. While common keywords like "nation" and "turkey" are shared with other clusters, it becomes evident that specific terms such as "government," "authority," "presidential system," "yes," "president," and "system" uniquely characterize Cluster 3. This thematic distinction sets Cluster 3 apart and is particularly noteworthy because it exclusively appears in speeches made prior to the 2017 constitutional referendum. The prevalence of these keywords in Cluster 3 speeches highlights the deliberate focus on topics related to the impending referendum, including the proposed transition to a presidential system and the associated changes in governance.

Cluster 4

When examining the speeches associated with Cluster 4, it becomes apparent that these conversations are not tied to a specific event or election period. Instead, these speeches appear to have been delivered during public gatherings held before the 2015, 2018, and 2023 elections. They predominantly feature keywords such as "nation," "country," "turkey," and "people," indicating a unifying and inclusive message conveyed to the public during these meetings.
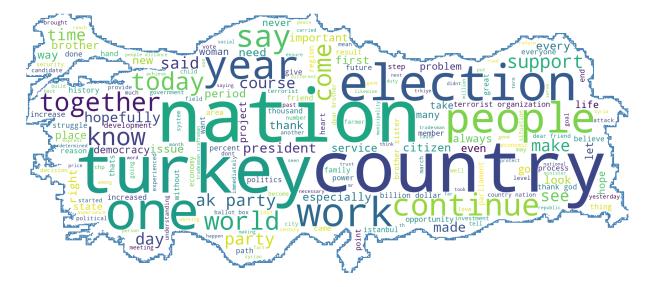
Cluster 5

On February 6, 2023, a catastrophic event unfolded in southern Türkiye near the northern Syrian border when a powerful magnitude 7.8 earthquake struck, followed by a second significant quake measuring 7.5 approximately nine hours later. These seismic events resulted in a tragic loss of life, with over 55,000 people losing their lives and another 100,000 individuals sustaining injuries in both Türkiye and Syria. In response to this disaster, Turkish President Recep Tayyip Erdoğan declared a three-month state of emergency in ten provinces of the country.

The disaster brought attention to critical issues such as the inadequate building inspections in high earthquake-risk areas and the shortcomings in implementing effective post-earthquake emergency aid strategies. Criticism also arose concerning the slow and inefficient delivery of essential aid. Consequently, the government placed a significant emphasis on addressing earthquake-related concerns in its speeches, which is evident in the five speeches delivered between February 7, 2023, and February 20, 2023. These speeches consistently align with Cluster 5, distinguishing them from other government communications, reflecting the government's ongoing commitment to addressing and rectifying the challenges posed by earthquakes in the region.
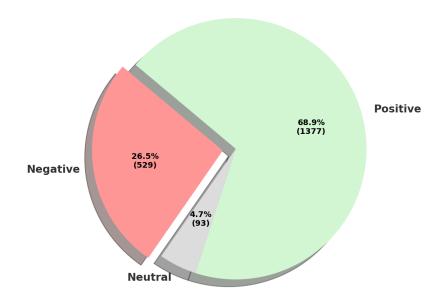
Cluster 6

When delving into the content within Cluster 6, several noteworthy patterns and correlations come to light. An intriguing observation is the consistent alignment of this cluster with speeches following victorious election outcomes. It becomes evident that speeches made immediately after significant electoral triumphs, including those following the 2014 presidential election, the 2015 1st Parliamentary elections, the 2017 referendum, the 2018 Parliamentary and Presidential Elections, as well as those preceding and succeeding the 2019 1st Local Elections and the 2023 Presidential Election, predominantly fall within Cluster 6. Given that these electoral victories were secured by Recep Tayyip Erdoğan and his party, a clear connection between this cluster and the nature of the speeches emerges.

Notably, this connection is reinforced by the prevalence of specific keywords such as "one," "nation," "country," "election," "work," "continue," and "together." These keywords resonate throughout Erdoğan's speeches, often centering around his prominent slogan, "One Nation, One Flag, One Homeland, One State." However, it is particularly in Cluster 6 that these related terms intensify, reflecting the post-election context and emphasizing the unity and national focus emphasized in these pivotal moments.

4.2. Sentiment analysis

The sentiment analysis, when contextualized with the clustering results of Recep Tayyip Erdoğan's speeches, offers a nuanced understanding of the emotional undertones in his public communications. The predominance of positive sentiment, with 1377 instances, aligns with Erdoğan's strategic communication style, particularly evident in clusters focusing on election campaigns and national unity (Clusters 2 and 6). These clusters, characterized by themes of patriotism and electoral success, naturally lend themselves to a more positive and unifying rhetoric.

In contrast, the negative sentiments, totaling 529 instances, are likely reflective of the challenges and crises addressed in various clusters. For instance, Cluster 0 (Istanbul and 2019 Local Elections) and Cluster 5 (Response to the 2023 Earthquake) encompass speeches delivered during periods of political loss and national tragedy, respectively. These contexts necessitate a tone that acknowledges difficulties and challenges, thereby explaining the negative sentiments observed.

The neutral sentiments, though the least frequent with 93 instances, suggest a balanced approach in certain speeches, possibly aligning with Cluster 4, where the speeches are not tied to specific events but focus on general themes of national interest.

Overall, the sentiment analysis, in conjunction with the clustering results, underscores Erdoğan's adaptability in tone and rhetoric, depending on the thematic context and the

audience's expectations. The positive sentiments in clusters related to electoral successes and patriotic messaging, the negative sentiments in speeches addressing political setbacks and national crises, and the neutral tones in more general communications, together paint a comprehensive picture of Erdoğan's strategic communication across different pivotal moments in Türkiye's recent history.

4.3. Topic modeling

The topic modeling results distinctly capture the varying themes in Erdoğan's speeches based on sentiment:

- The positive sentiment speeches focus on national unity, brotherhood, and political-social agendas. They seem to be aimed at rallying the public around common identities and political goals.
- The negative sentiment speeches, however, deal with challenges and concerns, ranging from national security threats to societal issues. They appear to be more focused on addressing problems and responding to challenges facing the nation.

This contrast in themes underlines Erdoğan's multifaceted approach in public speaking, alternating between positive, unifying messages and addressing more somber or challenging topics, depending on the context and purpose of the speech.

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights
0	nation	0.0165	istanbul	0.0082
1	country	0.0113	turkey	0.0077
2	brother	0.0092	brother	0.0077
3	people	0.0075	party	0.0070
4	turkey	0.0074	people	0.0065
5	one	0.0072	know	0.0062
6	say	0.0071	year	0.0051
7	like	0.0065	like	0.0050
8	dear	0.0063	work	0.0049
9	god	0.0059	one	0.0048

Topic 0 (Nation and Brotherhood):

This topic is intricately linked with Erdoğan's slogan "One Nation, One Flag, One Homeland, One State," reflecting a strong emphasis on national unity and identity. The recurrent usage of terms such as 'nation,' 'country,' 'brother,' and 'people' aligns closely with this slogan, encapsulating a sense of collective identity and solidarity among the Turkish people. The presence of emotionally charged and personal words like 'say,' 'like,' 'dear,' and 'god' further intensifies this narrative, creating a bond that transcends mere political rhetoric. Erdoğan's consistent referral to these themes, particularly through his slogan, suggests a strategic effort to unify the country under a singular national identity, which is a cornerstone of his political messaging. This unified narrative is pivotal in rallying support and maintaining a cohesive support base, especially in a diverse political landscape.

Topic 1 (Political and Social Themes):

This topic reflects a more pragmatic aspect of Erdoğan's speeches, focusing on the political and social landscape of Türkiye. The inclusion of terms like 'Istanbul,' 'turkey,' 'party,' and 'year' signifies discussions surrounding Erdoğan's political maneuvers, initiatives,

and his party's role in the socio-political fabric of Türkiye. The use of 'people' in this context suggests a broader societal focus, possibly indicating Erdoğan's efforts to address or comment on various societal issues and events. The presence of these themes underscores Erdoğan's approach to intertwining his political agenda with broader societal narratives, thereby presenting his party's initiatives as not just political decisions but as integral elements of Türkiye's social and cultural evolution.

Negative Sentiment LDA Model

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights
0	nation	0.0119	nation	0.0134
1	country	0.0098	country	0.0108
2	terrorist	0.0088	say	0.0106
3	organization	0.0085	one	0.0082
4	earthquake	0.0070	people	0.0079
5	turkey	0.0054	know	0.0073
6	people	0.0051	brother	0.0069
7	thousand	0.0047	state	0.0048
8	brother	0.0047	said	0.0048
9	citizen	0.0046	turkey	0.0047

Topic 0 (National Challenges):

This topic covers a range of issues, including the July 15 coup attempt, activities of terrorism groups, and the government's response to the 2023 earthquake. Key terms like 'nation,' 'country,' 'terrorist,' 'organization,' alongside references to the earthquake, reflect a narrative that focuses on various national crises. The inclusion of the earthquake in this topic indicates discussions around the government's emergency response and the challenges faced in the aftermath of the disaster. Erdoğan's speeches on these subjects are characterized by a tone of resilience and determination, emphasizing the government's role in addressing and overcoming these challenges.

Topic 1 (Societal Concerns):

In this topic, words like 'nation,' 'country,' 'say,' 'one,' and 'people' reappear, but with different accompanying words like 'know,' 'brother,' 'state,' and 'said.' This indicates a discourse around societal concerns, potentially involving dialogue or responses to public opinions or issues. The presence of 'state' might signify discussions on the role or actions of the government in relation to these concerns.

4.4. Word2Vec

Nation and Turkey

The application of Word2Vec, a powerful natural language processing tool, to analyze the political speeches of Recep Tayyip Erdoğan has yielded intriguing insights, particularly when examining the words "nation" and "turkey." The results from this analysis are profoundly reflective of Erdoğan's main slogan: "One Nation, One Flag, One Homeland, One State.".

```
• Nation

sgram_model.wv.most_similar("nation", topn=20)

[('state', 0.9662063121795654),
 ('one_flag', 0.9654416441917419),
 ('one_homeland', 0.9642665982246399),
 ('one', 0.9366958737373352),
 ('love', 0.926922619342804),
 ('democracy', 0.9211549758911133),
 ('brother', 0.9136342406272888),
 ('heart', 0.9084051251411438),
 ('struggle', 0.9071714878082275),
 ('turkish', 0.9069452881813049),
 ('member', 0.90458083152771),
 ('big_alive', 0.9001184701919556),
 ('believe', 0.9000380635261536),
 ('government', 0.8973979949951172),
 ('never', 0.8972773551940918),
 ('together', 0.8967845439910889),
 ('affection', 0.8959712982177734),
 ('yes', 0.8951825499534607),
 ('young_people', 0.895048201084137),
 ('friend', 0.8945595026016235)]
```

In the case of "nation," the most similar words include 'state,' 'one_flag,' 'one_homeland,' and 'one,' which resonate deeply with the components of Erdoğan's slogan. This suggests a strong emphasis on unity and patriotism in his speeches. Other words like

'love,' 'democracy,' 'brother,' and 'heart' indicate a rhetorical strategy that intertwines emotional appeal with nationalistic and democratic values. The presence of words such as 'struggle,' 'government,' and 'believe' also points towards a narrative of resilience and collective effort, a likely appeal to the public's sense of solidarity and determination.


```python
• Turkey

sgram_model.wv.most_similar("turkey", topn=20)

[('continue', 0.9803807735443115),
 ('hopefully', 0.9803512692451477),
 ('hope', 0.9762495160102844),
 ('goal', 0.9735961556434631),
 ('year', 0.9726955890655518),
 ('development', 0.9726792573928833),
 ('together', 0.9717786312103271),
 ('great', 0.9715016484260559),
 ('strong', 0.9709587097167969),
 ('build', 0.9687405228614807),
 ('power', 0.9681114554405212),
 ('economy', 0.9675856828689575),
 ('achieve', 0.966992199420929),
 ('different', 0.9650635123252869),
 ('th', 0.9650382995605469),
 ('stronger', 0.9649374485015869),
 ('country', 0.9649104475975037),
 ('step', 0.9648573398590088),
 ('period', 0.9646772742271423),
 ('experienced', 0.9638908505439758)]
```

For the word "turkey," the associated terms like 'continue,' 'hope,' 'goal,' 'development,' 'strong,' and 'build' reflect a forward-looking, aspirational tone in Erdoğan's discourse. These words suggest a focus on progress, strength, and national development, aligning with a vision of a unified, prosperous Türkiye. The emphasis on 'economy,' 'power,' and 'country' further underscores a narrative of national advancement and self-reliance.

The correlation between these Word2Vec results and Erdoğan's slogan is not coincidental. It underlines a deliberate rhetorical strategy designed to forge a strong national identity and a united vision for the country. This approach is evident in Erdoğan's consistent use of language that reinforces unity, strength, and progress – themes that are central to his political narrative and resonate deeply with the slogan "One Nation, One Flag, One Homeland, One State."

FETO and Terrorist

The Word2Vec analysis of Recep Tayyip Erdoğan's speeches, focusing on the words "feto" and "terrorist," reveals a deep and intricate relationship between the rhetoric used by Erdoğan and the significant events in Türkiye, particularly the July 15, 2016, coup attempt and the broader context of terrorism.

```
• FETO

sgram_model.wv.most_similar("feto", topn=20)

[('separatist', 0.9974353313446045),
 ('fethullahist', 0.9957336187362671),
 ('supporter', 0.9956838488578796),
 ('treason', 0.9956491589546204),
 ('called', 0.995609700679779),
 ('control', 0.995357096195221),
 ('betrayal', 0.9952726364135742),
 ('pkk', 0.9952429533004761),
 ('gang', 0.9952167868614197),
 ('attempted', 0.9951922297477722),
 ('daesh', 0.9949925541877747),
 ('instead', 0.9949464797973633),
 ('openly', 0.9948726296424866),
 ('organization', 0.9948456287384033),
 ('using', 0.9947811365127563),
 ('rely', 0.9947754144668579),
 ('armed_force', 0.9947474002838135),
 ('lifestyle', 0.9947146773338318),
 ('qandil', 0.9946256875991821),
 ('majority', 0.9945505261421204)]
```

For the term "feto," the analysis shows strong associations with words like 'separatist,' 'fethullahist,' 'supporter,' 'treason,' 'betrayal,' and 'gang.' These associations are highly indicative of Erdoğan's framing of the Fethullah Gülen Terrorist Organization (often referred to as FETÖ) in the context of the 2016 coup attempt. The strong connection with words like 'treason' and 'betrayal' underscores the portrayal of FETÖ as a clandestine and treacherous organization, undermining the state. Erdoğan's speeches seem to systematically associate FETÖ with betrayal and treachery, tying it to the failed coup attempt. The presence of words like 'pkk' and 'daesh' in the same context further expands this narrative, linking FETÖ to wider themes of terrorism and security threats.

- Terrorist

```
sgram_model.wv.most_similar("terrorist", topn=20)
```

```
[('daesh', 0.9946590065956116),
 ('terrorism', 0.9943891763687134),
 ('threat', 0.9941885471343994),
 ('syria', 0.994113028049469),
 ('structure', 0.9940972328186035),
 ('civilization', 0.9937019944190979),
 ('tried', 0.9936355948448181),
 ('value', 0.9933659434318542),
 ('culture', 0.9932531714439392),
 ('treason', 0.9930546879768372),
 ('muslim', 0.9927589297294617),
 ('current', 0.9926144480705261),
 ('even_though', 0.9923701286315918),
 ('loss', 0.9923328161239624),
 ('oil', 0.9922627210617065),
 ('try', 0.9922107458114624),
 ('global', 0.9921724200248718),
 ('anatolian', 0.9921254515647888),
 ('vision', 0.9921090602874756),
 ('space', 0.9921038150787354)]
```

When examining the word "terrorist," there is a clear linkage to terms such as 'daesh,' 'terrorism,' 'threat,' 'syria,' and 'civilization.' This reflects the broader scope of Erdoğan's discourse on terrorism, extending beyond the domestic sphere to include international terrorism, with a particular focus on groups like Daesh (ISIS) and the conflicts in Syria. The association with words like 'civilization,' 'culture,' and 'muslim' suggests an attempt to frame terrorism as a threat not only to national security but also to cultural and religious values.

Earthquake and Disaster

The Word2Vec analysis of Recep Tayyip Erdoğan's speeches, particularly with a focus on the terms "earthquake" and "disaster," provides a revealing perspective on the linguistic framing of the catastrophic 2023 earthquake in Türkiye.

- Earthquake

```
sgram_model.wv.most_similar("earthquake", topn=20)
```

```
[('area', 0.9757002592086792),
 ('caused', 0.9717267751693726),
 ('building', 0.9698222279548645),
 ('production', 0.9669913053512573),
 ('student', 0.9669901132583618),
 ('destruction', 0.9663795232772827),
 ('disaster', 0.9640430808067322),
 ('public', 0.9639953970909119),
 ('aid', 0.9639552235603333),
 ('personnel', 0.9637019038200378),
 ('approximately', 0.9632876515388489),
 ('increase', 0.9629559516906738),
 ('increased', 0.9628385901451111),
 ('unit', 0.9618889093399048),
 ('billion_lira', 0.9611796736717224),
 ('ton', 0.9609955549240112),
 ('capacity', 0.9609662294387817),
 ('dollar', 0.96079021692276),
 ('employment', 0.9606766700744629),
 ('facility', 0.9604082107543945)]
```

For the word "earthquake," the associated terms such as 'area,' 'caused,' 'building,' 'destruction,' 'disaster,' 'public,' 'aid,' and 'personnel' offer a comprehensive view of the earthquake's impact and the subsequent response. These words underscore the widespread destruction and the extensive efforts in terms of aid and recovery.

- Disaster

```
sgram_model.wv.most_similar("disaster", topn=20)
```

```
[('destruction', 0.9940009713172913),
 ('occurred', 0.9931188821792603),
 ('magnitude', 0.9929953813552856),
 ('loss', 0.991802990436554),
 ('due', 0.9916819334030151),
 ('february', 0.991590142250061),
 ('wound', 0.9914592504501343),
 ('destroyed', 0.9914546012878418),
 ('caused', 0.9913880825042725),
 ('condition', 0.9913578033447266),
 ('major', 0.9912469387054443),
 ('completely', 0.9911360144615173),
 ('effect', 0.9910407662391663),
 ('progress', 0.9909842610359192),
 ('damage', 0.9908450245857239),
 ('serious', 0.9908163547515869),
 ('crisis', 0.9907767176628113),
 ('victim', 0.9907212853431702),
 ('experienced', 0.9906888604164124),
 ('past_year', 0.990610659122467)]
```

In the context of "disaster," the results show a strong correlation with words like 'destruction,' 'magnitude,' 'loss,' 'wound,' 'destroyed,' 'crisis,' and 'victim,' painting a vivid

picture of the scale and human toll of the earthquake. The word 'february' directly ties to the time of the earthquake, emphasizing its temporal significance. The terms 'serious,' 'crisis,' 'damage,' and 'condition' highlight the severity and the urgent need for response and recovery efforts. This is in line with the government's declaration of the highest level of emergency and the international assistance received.

The connection between these Word2Vec results and the devastating 2023 earthquake in Türkiye is profound. Erdoğan's speeches reflect a multi-faceted approach to the disaster, addressing both the immediate humanitarian crisis and the longer-term economic and infrastructural challenges. The emphasis on 'building,' 'aid,' and 'personnel' points to the immediate response efforts, while terms like 'billion_lira,' 'dollar,' and 'employment' indicate a recognition of the economic fallout and the need for substantial recovery efforts.

5. Conclusion

The exploration into Recep Tayyip Erdoğan's political speeches through advanced text mining has yielded comprehensive insights into his rhetorical strategies and thematic focuses. Through the careful application of clustering, sentiment analysis, topic modeling, and Word2Vec analysis, our three hypotheses regarding Recep Tayyip Erdoğan's rhetorical strategies and thematic focuses have been meticulously investigated and confirmed.

The first hypothesis regarding Erdoğan's nationalistic sentiments finds substantial support in our analysis. The analysis of Erdoğan's speeches, particularly within Clusters 2, 3, 4 and 6, revealed a consistent emphasis on nationalistic sentiments. Beyond the clustering of speeches, the Word2Vec results and topic modeling provide a more nuanced view. The Word2Vec analysis, particularly regarding the word "nation," revealed strong associations with terms like 'state,' 'one_flag,' 'one_homeland,' and 'unity.' This linguistic pattern aligns seamlessly with Erdoğan's slogan "One Nation, One Flag, One Homeland, One State," underscoring a strategic emphasis on national unity and identity. Similarly, topic modeling accentuated this theme, with topics around 'nation and brotherhood' and 'political and social themes' echoing a consistent narrative of national pride and collective identity. These findings not only confirm the hypothesis but also highlight Erdoğan's skillful weaving of nationalistic rhetoric throughout his political discourse, aiming to unify the populace under a shared national narrative.

In examining the second hypothesis about Erdoğan's rhetorical shifts during elections and crises, the clustering results from Cluster 1 and Cluster 5 are particularly telling. These clusters starkly differ from others, encapsulating speeches that respond to specific and high-stake situations like the 2016 coup attempt and the 2023 earthquake. These speeches are distinct in their thematic focus, with a heightened emphasis on resilience, national unity, and governmental response in times of crisis. This thematic deviation from other clusters, such as those focusing on elections or general political messaging, is indicative of Erdoğan's adaptive rhetorical strategy, modulating his speech according to the prevailing political context and the nation's sentiment.

The third hypothesis, concerning the dynamics of sentiment in Erdoğan's speeches, is corroborated through the detailed sentiment analysis. Positive sentiments, prevalent in 1377 instances, were predominantly observed in speeches relating to electoral successes and moments of national celebration. These instances reflect a strategic use of uplifting, unifying rhetoric, particularly aimed at rallying support and fostering a positive public perception. In contrast, the 529 instances of negative sentiment corresponded largely to speeches addressing national crises or political setbacks. These speeches often conveyed a tone of solemnity and determination, acknowledging the challenges while endeavoring to instill a sense of resilience.This dichotomy in sentiment highlights the strategic adaptation in Erdoğan's speeches. The contrast between positive and negative tones in different contexts suggests a calculated approach to aligning his rhetoric with the prevailing national mood and political circumstances.

In conclusion, the findings of this study support the initial hypotheses, offering a nuanced perspective on Erdoğan's rhetorical methods. The analysis, grounded in advanced text mining techniques, sheds light on Erdoğan's speech patterns, indicating a pragmatic approach to public communication. His speeches, varying in tone and theme, seem to reflect a response to the diverse demands and challenges of the political landscape in Türkiye.

6. Appendix

For further reference and a detailed overview of the methodologies, code, and data used in this study, please visit the following GitHub repository. This repository contains all the scripts, datasets, and analysis files used in our study. It provides a comprehensive

resource for understanding the text mining techniques applied, including clustering, sentiment analysis, topic modeling, and Word2Vec analysis.

Repository Link: https://github.com/mrcljns/Text-Mining-Projects