To cite: Oğlakçı, B., Uzun, A. (2024). Social Media Data Mining as a Novel Approach to Electoral Geography, *Coğrafî Bilimler Dergisi/ Turkish Journal of Geographical Sciences*, 22 (1), 137-156, doi: 10.33688/aucbd.1407317

Coğrafi Bilimler Dergisi
Turkish Journal of Geographical Sciences
e-ISSN:1308-9765

Social Media Data Mining as a Novel Approach to Electoral Geography

Seçim coğrafyasına yeni bir yaklaşım olarak sosyal medya veri madenciliği

Burak Oğlakcı*[a], Alper Uzun[b]

Makale Bilgisi

Araştırma Makalesi

DOI: 10.33688/aucbd.1407317

Makale Geçmişi:
Geliş: 20.12.2023
Kabul: 11.03.2024

Anahtar Kelimeler:
Seçim coğrafyası
Bağlam etkisi
Doğal dil işleme
Sosyal medya
X (Twitter)

Öz

Sosyal medyada yer alan bireysel söylemler ve tutumlar, sahip oldukları mekânsal-zamansal bilgilerle birlikte seçim coğrafyası için kullanıldığında seçmen davranışları ve oy tercihleri hakkında önemli ipuçları sağlayabilmektedir. Sosyal medya verilerini kullanan seçim odaklı farklı çalışmalar olsa da bu çalışmalar oyların ardındaki nedenleri mekânsal olarak anlamaktan uzak kalmıştır. Bu araştırma, seçim coğrafyası alanı için yeni sayılabilecek veri ve yöntemleri kullanarak seçmen davranışı ve oy vermedeki farklılıkları anlamayı, oy verme üzerindeki bağlamsal etkileri ortaya çıkarmayı amaçlamaktadır. Araştırmada X (Twitter) üzerinde Yenimahalle (Türkiye) ilçesinden paylaşılan seçim odaklı tweetleri elde ediyor, duygu durumlarını belirliyor, tweetlerdeki baskın konuları tespit ediyor ve adayların X (Twitter) üzerindeki popülerlik oranları ile gerçek oy oranları arasındaki ilişkiyi ortaya koyuyoruz. Sonuç olarak, X (Twitter) üzerinde kullanıcılar tarafından oluşturulan mekân-zamansal veriler, seçmen davranışı/katılımı ve oy vermedeki farklılıkları anlamanın yanı sıra oy verme üzerindeki bağlam etkilerinin ortaya çıkarılmasında önemli bir yer tutmaktadır.

Article Info

Research Article

DOI: 10.33688/aucbd.1407317

Article History:
Received: 20.12.2023
Accepted: 11.03.2024

Keywords:
Electoral geography
Context effects
Natural language processing
Social media
X (Twitter)

Abstract

When individual discourses and attitudes on social media are used for electoral geography with their spatio-temporal information, they can provide essential clues about voter behavior and voting preferences. Although there are different election-based studies using data from social media, these studies have remained far from spatial understanding of the reasons behind the votes. This research aims to understand the differences in voter behavior and voting, along with revealing the context effects on voting by using this new data and methods in the field of electoral geography. We obtain the election-oriented tweets on X (Twitter) posted from Yenimahalle (Turkey), determine sentiment states, identify the dominant topics in the tweets, and reveal the relationship between the popularity rates of the candidates on X (Twitter) and the actual vote rates. Consequently, user-generated spatio-temporal data on X (Twitter) is essential in understanding the differences in voter behavior/participation and voting, as well as in revealing context effects on voting.

***Sorumlu Yazar/Corresponding Author:** burak.oglakci@cbu.edu.tr
[a] Manisa Celal Bayar Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, Coğrafya Bölümü, Manisa/Türkiye, https://orcid.org/0000-0003-4559-4108
[b] Balıkesir Üniversitesi, Fen-Edebiyat Fakültesi, Coğrafya Bölümü, Balıkesir/Türkiye, https://orcid.org/0000-0002-1304-1683

1. Introduction

With the advent of Web 2.0, the traditional one-sided understanding of information in Web 1.0 has transitioned to a new dimension. Users are no longer just consumers of the Internet; they have become active participants and stakeholders in the system. This is due to the development of social media platforms, wikis, blogs, forums, RSS and more, where users can produce their content and communicate with others.

The aforementioned platforms, networks, and apps have become spaces where spatio-temporal information is collected and accessed in line with the activities carried out by users for various purposes. Behind this potential is the fact that users leave various pieces of information about themselves and the environment they live in, including their location, on these platforms, networks, and apps (Bermingham and Smeaton, 2011; Goodchild, 2007; Harrison and Barthel, 2009; Murugesan, 2007).

Based on recent developments, there has been a noticeable rise in the usage of social media platforms like X (Twitter), Facebook, Instagram and so on. Alongside this increase, there has also been a growing significance attributed to the discussions, ideas, and attitudes within these platforms. As a result, there has been an increase in research that focuses on the content generated by users on these platforms. An important aspect to consider is that the spatio-temporal information of the shared content also offers valuable data potential for geographical research. When we examine communication and interaction on these platforms within the context of electoral geography, they offer important insights into voter behavior, voting preferences, and the influencing factors behind these preferences. In the literature, there are electoral studies using data on social media platforms, and X (Twitter) is often preferred by researchers. However, to date, there have been very few geographical studies that address this issue.

X (Twitter) is one of the Location-Based Social Networks (LBSN) that contains a large amount of spatio-temporal information and allows users to share posts, called “tweets,” limited to 280-character-plain text or with images and videos. X (Twitter) also offers its users the opportunity to add real-world location information to tweet content since 2009. Thanks to these opportunities, the large amount of spatio-temporal data accumulated on X (Twitter) is also being used in many other research areas (Oğlakcı and Uzun, 2021). Although previous electoral studies using data on X (Twitter) have addressed the estimation of vote rate, and users’ behavior on the social network during the election period, these studies have tended to remain far from spatial understanding of the reasons behind the votes (arising from space, place, and socio-political context). This points to an opportunity for future research in electoral geography that can be conducted by processing and utilizing the large amount of user-generated spatio-temporal data on X (Twitter).

The research presented in this article focuses on uncovering the reasons behind vote preferences, making vote predictions, and the contributions that can be provided to the field of electoral geography by using data from X (Twitter) and the social media mining method. We demonstrate the potential of this approach by examining election-oriented posts on X (Twitter) shared from Yenimahalle (Ankara, Turkey), determining sentiment states, identifying dominant topics (context effects) in these posts, and revealing the relationship between the popularity rates of the candidates on X (Twitter) and the actual vote rates.

1.1. Electoral Geography and Social Media (Data) Mining

A subfield electoral geography examines the relationship between space, place, and elections. It serves as a connection between human geography and political science (Pattie and Johnston, 2009). Democracy and electoral geography are intimately related in Shelley's (2006) study. According to Shelley, this field examines data that has been geographically separated to understand the differences in voting patterns among various geographic areas. According to McPhail (1971) and Reynolds (1990), research in electoral geography typically revolves around three main topics: the geography of voting, geographical influences on voting, and the geography of representation.

Studies on the effects of technological developments on voter behavior and voting are included in the literature. For example, Books and Prysby (1991) in their study addressed the impact of information flow through the media on the relationship between local context and voting behavior. According to Agnew (1996), the presence and availability of communication technology are determining factors for the level of interaction within a given space.

However, more recent studies on this topic, such as Adams (2015) and Shin (2015) have noted a growing interest in studying the impact of technology and social media on electoral behavior and civic participation. Once again, Shin (2015) directs attention to the subject and underscores the need to fully comprehend the connection between these technologies and electoral geography. Similarly, Temple (2023) follows two cases on electoral propaganda and digital developments in the UK and US elections, focusing on what the use of digital technology and the increasingly digitized electoral environment will bring for electoral geography and, perhaps, the possibility of ending some of its approaches.

Social media (data) mining carries great potential for understanding voter behavior and voting preferences and for revealing contextual effects on these preferences. Data mining, also known as knowledge discovery in data (KDD) (IBM, 2021), can be defined as the process of discovering new, useful, interesting information from large amounts of existing data, as a result of the extensive accessibility of large amounts of data and the emergence of the need to transform such data into useful information (Han and Kamber, 2001). Social media platforms, which have an enormous amount of user-generated content as a result of their use for various purposes, are a suitable area for data mining (Gundecha and Liu, 2012). Social media (data) mining, which involves mining existing data together with social relations, can be described as "the process of representing, analyzing, and extracting meaningful patterns from data in social media" (Zafarani et al., 2014: 21).

Behind the potential of using the data gathered from social media platforms in the field of electoral geography, there are various advantages provided by these platforms. It is possible to list these advantages by taking inspiration from the studies of Seymour (2001), Kozinets (2002), Markham (2004), and Convery and Cox (2012), Yıldırım and Şimşek (2016) mentioning the convenience and availability of social media data for researchers and research. Social media platforms provide the opportunity to access large amounts of data, fields and connections that cannot be reached with traditional research methods (surveys, interviews, etc.), as well as providing control and flexibility in time, space, volunteering, and communication processes. However, the platforms offer privacy and freedom from physical constraints, allowing individuals and groups to express their views more comfortably, sincerely, and freely.

In line with aforementioned advantages, there are various electoral studies that specifically target user data on X (Twitter) and use appropriate methods (Belcastro et al., 2020; Chauhan et al., 2023; Jaidka et al., 2018; Khan et al., 2023; Marozzo and Bessi, 2018; Oikonomou and Tjortjis, 2018; Ramteke et al., 2016; Soler et al., 2012; Song et al., 2014; Tumasjan et al., 2011; Wang and Gan, 2018). However, existing research has tended to be more limited in terms of exploring the reasons behind voting preferences and addressing contextual effects. On the other hand, there are also limitations in determining the place of these approaches in electoral geography. In conclusion, the use of data on relevant platforms and social media data mining method constitutes an important emerging approach in the field of electoral geography, and also one that deserves greater scrutiny.

1.2. Local Elections in Turkey, 2019 Local Elections and Yenimahalle

There are various studies on electoral dynamics in Turkey and party preferences of Turkish voters from a geographical perspective. For example, Çarkoğlu and Avcı (2002) categorize political parties in Turkey to three groups (center-right and center-left, Islamist-nationalist), reveal breaks and deviations for the parties in the historical process, and identify geographical regions that have come to be associated with different ideological orientations. Çarkoğlu and Hinich (2006) adopt the framework of spatial voting theory to address the pro-Islamic bias in the periphery against the secular center and focused on the reasons and changes in voting preferences. Wuthrich (2015) examines Turkish general elections and observes tendencies towards pragmatism in the voting behavior of Turkish voters. Özen and Kalkan (2017) focus on the geographical patterns of party votes and competition in Turkey using spatial models.

The literature on Turkish elections is vast, but there are some notable gaps within it. For example, despite some researchers' attention to local political factors and dynamics, there is notably less interest in examining local elections in Turkey than general elections. Concentrating only on candidate and party performance in research on local elections also limits the understanding of local dynamics affecting elections (Bayraktar, 2011; Kamalak, 2013). From a broader perspective, it is possible to evaluate local elections as an intersection point, shaped by both national and local influences. Some studies accordingly examine the determining role of local factors on voter preference (Kamalak, 2013).

Our study differs from previous research in three important ways. Firstly, we use more diverse data and obtain it in a different manner. Most existing research on Turkish elections is based solely on official vote results, whereas our study incorporates posts from X (Twitter) during the election period as well as the official vote results. On the other hand, spatial units of different sizes may contain different clues and suggest different patterns of voting preferences. Aggregated results from larger spatial units may cause local dynamics to be overlooked. For this reason, we focus on a smaller area of analysis (a county) in our research process. Finally, our study traces the local and national level factors that influence vote preferences. This is because the intense electoral communication and information flow with spatial reference that takes place in the social network includes various issues, factors, and attitudes.

1.2.1. The Intersection of Local Governments and National Politics: Local Elections in Turkey

Local elections in Turkey were first held in 1930 and are planned to be held every five years. There has been some variation in that schedule, based on the vicissitudes of historical circumstances. In each election, citizens elect their administrators to carry out local services in their village, neighborhood, town, county, or province. Although the functioning of democratic processes in Turkey was disrupted by military coups in 1960 and 1980, local elections have been held regularly every five years since 1984.

In the 1990s, local elections began to take on a different dimension for parties and candidates, and these transformations inform geographical understandings of Turkish politics today (Bekaroğlu and Kaya Osmanbaşoğlu, 2021). This is because these elections paved the way for parties and candidates who received high votes, especially from metropolitan cities such as Istanbul, Ankara, and Izmir, to gain an important place in national politics (İncioğlu, 2002). The last local election in Turkey was held on March 31, 2019. In this election, basically two alliances were active, and the election was shaped by bloc politics. The first alliance, the People's Alliance, consisted of the AK Party (Justice and Development Party) and the MHP (Nationalist Movement Party), while the second, the Nation Alliance, consisted of the CHP (Republican People's Party) and the IYI Party.

The March 31, 2019 Election, in addition to being an election in which alliances are effective, is the first local election after the transition from the parliamentary system to the Presidential Government System on April 16, 2017. The alliances formed before the election had a significant impact on the atmosphere of this election because pragmatism rather than ideological unity was effective in forming the alliances. On the other hand, alliances changed the party blocs (right-left) and increased the vote transition between the blocs (Miş and Duran, 2019). As a matter of fact, while transitions within the blocs are common in Turkey, the transition between the blocs has historically been limited (Kalaycıoğlu, 1999). Finally, in the March 31, 2019 election the People's Alliance lost the elections in Turkey's metropolitan cities such as Ankara and Istanbul, and the Nation Alliance candidates achieved control of metropolitan municipality administration. Thus, the supremacy of the AK Party, which won the elections in Ankara with the candidates it nominated in the local elections in 1999 and later, came to an end. Similarly, ongoing dominance in Istanbul since 2004 has come to an end. The March 31, 2019 election are important in the history of Turkey's local elections, with the establishment of two major alliances, the expansion of the transition between party blocs, change of leadership in administration of two metropolitan cities, and the success of the Nation Alliance candidates in gaining significant vote rates.

1.2.2. Why Yenimahalle?

Yenimahalle is a county in Ankara, the capital city of Turkey. When selecting this county as the study area, factors such as political preference status, population, and the presence, quantity, and quality of posts with spatio-temporal references that make up the research data were taken into consideration. During the 141-day data collection period for the March 31, 2019 Election, Çankaya was the county with the highest number of posts among the counties in Ankara. Yenimahalle came in second place. However, Çankaya was not chosen as a research area due to its political and cultural composition. This is because voters in Çankaya tend to vote based on their loyalty to the CHP (Republican People's Party).

The voting preferences of this county in local elections in 1989-1994 (SHP, Social Democrat People's Party - a party established after CHP was closed during the military coup in 1980), 1999 (CHP), 2004 (CHP), 2009 (CHP), 2014 (CHP), and 2019 (CHP) support this consideration. In the 2019 election, CHP candidates received 76.8% of the votes for the Ankara Metropolitan Municipality (AMM) and 73.4% for the Çankaya Municipality from the Çankaya county. On the other hand, due to their loyalty to the CHP, voters in Çankaya "act with the same sensitivity and attitude in local elections as in general elections" (Şahin et al., 2014: 167). Based on this information, the possibility that local problems and local political dynamics are being ignored in the voting behavior of Çankaya county increases.

When analyzing Yenimahalle in terms of politics, it can be observed that individuals with diverse political and cultural affiliations live together in the county, and this is reflected in the political composition of the county. Additionally, parties from both the right and left sides of the political spectrum have played an active role in county administration during certain periods (Şahin and Gözcü, 2017). This situation is also reflected in the preferences of the county in AMM elections. Finally, the tweets about the AMM election from Yenimahalle county cover various topics and factors. These issues and factors provide an important opportunity to understand the reasons behind electoral behavior and voting preferences.

2. Data and Method

The method of the research was designed to reach tweets about the election and related elements from the county of Yenimahalle during this election period, as well as to make sense of these tweets within the scope of electoral geography. In the study, election posts/tweets containing spatio-temporal information were used as data. To analyze and interpret the data, the following steps were taken: data preprocessing, application of predefined keywords, sentiment analysis, topic modeling, and calculation of candidate representation and popularity. Figure 1 illustrates the steps followed in the data collection and analysis process.

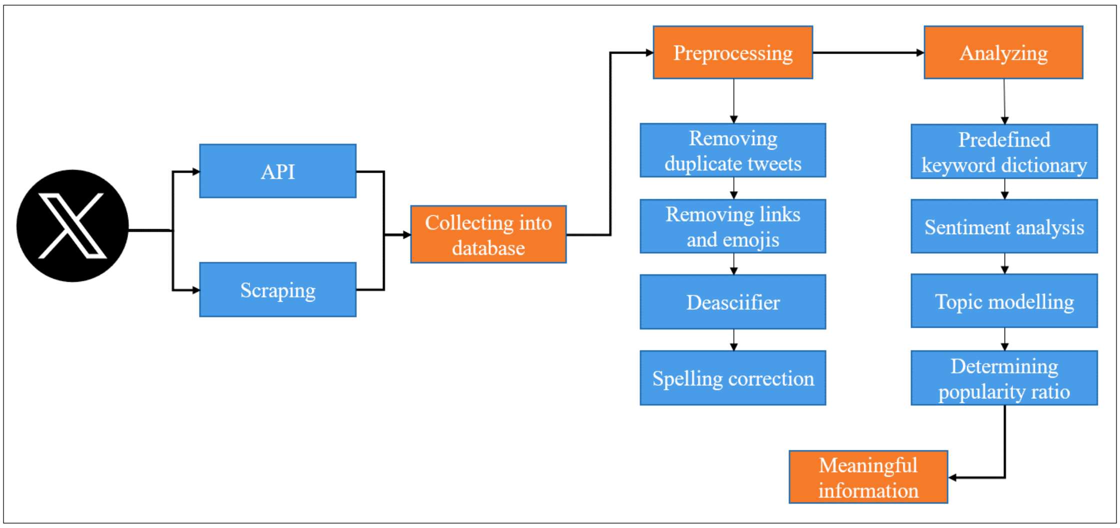

Figure 1. Steps followed in the process of data collection and analysis

The techniques followed in the data collection and analysis steps have numerous advantages. As mentioned in the previous subsection (1.1.), these advantages can be listed in a diverse manner, drawing inspiration from Seymour (2001), Kozinets (2002), Markham (2004), Convery and Cox (2012), Ceron et al. (2014), Yıldırım and Şimşek (2016), Snelson (2016), Olshannikova et al. (2017), Jaidka et al. (2018), and Liu et al. (2021).

Firstly, data on social media platforms and the techniques developed to process it provide an alternative to and/or complement traditional data collection methods such as surveys and interviews. Additionally, these techniques enable access to and rapid processing of large amounts of data, known as big social data, which is generated in a more relaxed, candid, and voluntary manner. Finally, the advantages of these platforms and techniques include time-space flexibility and the ability to replicate phases.

2.1. Collecting Data from X (Twitter)

In the research, data was collected for 141 days from December 1, 2018, when the candidates were announced for the March 31, 2019 Local Elections, to April 20, 2019. Since the election communication continued after the voting day, data collection was continued. Data was obtained from X (Twitter) in two ways. The first method is the real-time collection and storage of tweets located within the specified coordinates (32°28'E, 39°39'N), (33°12'E, 40°13'N) using the X (Twitter) API bounding box query[1]. The bounding box used and related information are shown in Figure 2.

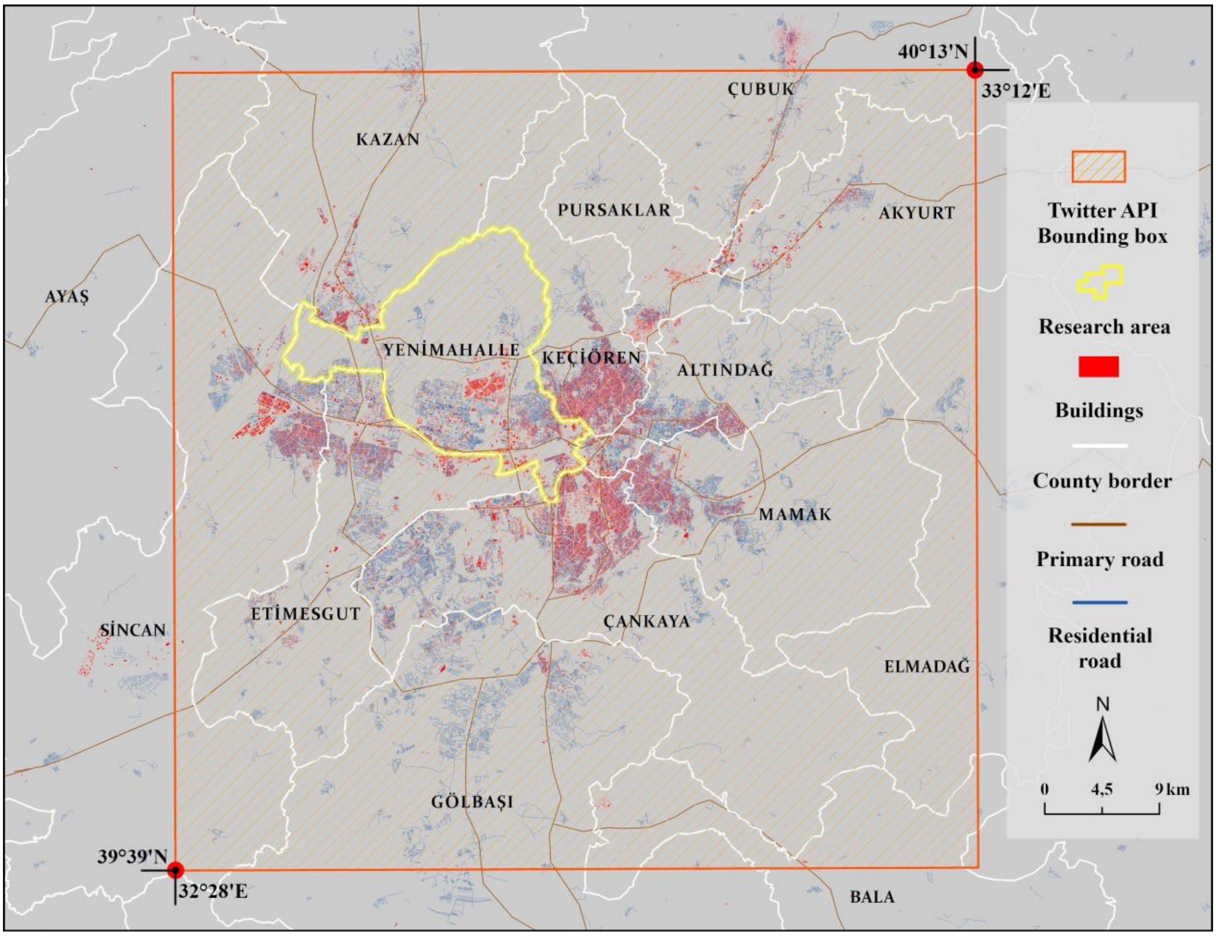

Figure 2. Information on the X (Twitter) API bounding box query

The other method is the process called web scraping or web harvesting. “Web scraping is an automated process that involves extracting and organizing specific data from websites” (GeeksforGeeks, 2023). In this method, the geolocation filters (point radius and place id) and the advanced search section of X (Twitter) were used. We ran a point radius filter by entering the latitude, longitude, and radius information for a particular area, such as 39.956389,32.803333,1.7km/mi. X (Twitter) provides the capability to access tweets that remain in the field, created in accordance with this information. All the parameters used for Yenimahalle are given in Table 1.

Table 1. Parameters used in point radius filter for Yenimahalle

Coordinates (lat, long)	Radius (km)
39.939167, 32.824722	0.65
39.946111, 32.775278	1
39.956389, 32.803333	1.7
39.995278, 32.734444	5

“X (Twitter) social network assigns a distinct identification string to every location or area such as a building, structure, neighborhood, county, province, or country” (Tweet Location Metadata, 2023). When an area or location is queried using its unique string, X (Twitter) provides access to tweets containing spatial information shared from that area or location. For example, the string "6e9cef2a932796f9" corresponds to the location of Yenimahalle, Turkey on X (Twitter). When a request using this string is run, X (Twitter) returns tweets with location information associated with Yenimahalle. Below is a sample post, and location information reached with this request is shown (Figure 3).

Figure 3. A Sample content reached by place id filter

Finally, tweets that mentioned location information were examined by entering word pairs (e.g., Yenimahalle and vote) in the advanced search section of X (Twitter). The aim here is to reach the tweets created by the people living in Yenimahalle by both talking about the elections and mentioning the location information. Other words used with Yenimahalle are *"election," "vote," "candidate," and "ballot box."* An example tweet with location information reached with word pairs is given below.

(Turkish) Bizden oy almayı düşünenler, Kardelen mahallesi, Yenimahalle'deki kötü durumu bir gelip görseler iyi olur. Bakalım oy alabilecekler mi?

(English) Those who are seeking our votes should come and witness the poor condition of Kardelen Street, Yenimahalle. Let's see if they can earn our votes.

2.2. Data Preprocessing

To increase the success rate of the processes applied in the method section, preprocessing is required on the data obtained from X (Twitter). This is because the tweets are mostly not created in accordance with the spelling rules. Another problem is the presence of large amounts of meaningless and mass posts. For this reason, various editing and cleaning processes were applied to the data using Natural Language Processing (NLP) libraries. The following steps were applied to the tweets:

- *Fake, meaningless, and repetitive tweets were excluded from the dataset.*
- *The specified links (https://twitter.com/, pic.twitter.com/, https://t.co/, https//www.swarmapp.com/c/, https://www.instagram.com, https://www.instagram.com/p/) and emojis have been cleared from the tweet text.*
- *Deasciifier process was applied to the words in which Turkish-specific characters (ç, ğ, ş, ü, ö, i, ı) are not used.*
- *A spelling correction process has been applied to correct misspelled words.*

2.3. Identifying Election-Oriented Posts/Tweets

A dictionary was created to identify the election-oriented posts among all the collected posts and to determine which candidate, party, and alliance these posts are for. There are studies in the literature that use the predefined keyword dictionary approach to identify the target entities mentioned in the posts (Cantey, 2013; Sakaki et al., 2010; Scanfeld et al., 2010). Similarly, in this research, the keyword dictionary approach was used to determine the target entities of the posts. While creating the keyword dictionary, alliances, elected officials, parties, and candidates were taken into consideration. The dictionary can be accessed at (*https://github.com/burakoglakci/Predefined-keyword-dictionary-for-Turkey-2019-Local-Election*). Before implementation, all characters were converted to lowercase using the string lower function to avoid case-sensitive errors. Then, the data kept in JSON format was processed using the Tableau program, which is an interactive tool for data analysis and visualization.

2.4. Statistical Significance of the Data

In line with the spatial approach adopted in the study, all the tweets included in the analysis have location information. This location information was associated with Yenimahalle (the research area) using the tweet location, profile location, and the mentioned location. Additionally, in all analyzed tweets, the lang (language) field is equal to "tr" (Turkish).

On the other hand, the Pearson Correlation value was calculated to determine the statistical significance of the data. The correlation coefficient for our data is 0.9, indicating a strong correlation between the data. When applying this analysis, tweets with location information posted from 6 central counties of Ankara, including Yenimahalle (visible in Fig. 2), within a 141-day period, and those election-oriented were used. In the 141-day period, the comparison of the total number of tweets with location information (Total - T) and the number of election-oriented tweets (Election - E) among them by county is as follows: Yenimahalle (T: 65,635 – E: 5,755), Çankaya (T: 160,460 – E: 12,957), Etimesgut (T: 47,372– E: 5,103), Keçiören (T: 44,124 – E: 4,703), Altındağ (T: 25,551 – E: 1,347), Mamak (T: 24,966– E: 2,949).

Finally, the suitability of the 5,755 election-oriented tweets for Yenimahalle in terms of sample size was also calculated. Using the sample size (5,755), Yenimahalle population (663,580), and confidence level (95%), the margin of error for the sample size was calculated as 1.29%. This low margin of error indicates the reliability of the results.

2.5. Sentiment Analysis of Election Posts

In this step, to analyze the sentiment of election posts, sentence-level and model-based sentiment analysis were utilized. The analysis employed the model developed by Köksal (2020) using Google BERT (Bidirectional Encoder Representations from Transformers) which accurately represents a sentence by extracting a numerical vector. This model was chosen because it is a new model for sentiment analysis in Turkish texts and has the ability to effectively detect complex situations.

While creating the training data of the model, 1,840 election posts (523 negative, 672 neutral and 645 positive), which represent the existing data very well and whose sentiment state was determined manually, were used. After these data were used as training data, the model was tested on 862 posts that it had not seen before and whose sentiment state was determined manually. After the training, the model achieved success in the range of 83-85% on the test data. The success rates in the model achieved with two different run results are shown below (Figure 4).

	precision	recall	f1-score	support
0	0.88	0.89	0.88	258
1	0.81	0.87	0.84	293
2	0.88	0.80	0.84	311
accuracy			0.85	862
macro avg	0.85	0.85	0.85	862
weighted avg	0.85	0.85	0.85	862

Figure 4. (%) Success rates obtained in the model (0: negative, 1: neutral, 2: positive)

2.6. Topic Modeling of Election Posts

With sentiment analysis, the sentiment states of the posts were determined as positive, neutral, and negative. Unlike the neutral states, positive and negative posts contain important information about the vote preferences of the voters and the factors affecting these preferences. For this reason, topic modeling was used to understand the background of the election posts and to make inferences about their general themes and topics. The Latent Dirichlet Allocation (LDA) algorithm was preferred when applying topic modeling.

A machine learning technique to learn about a text's semantic structure is called topic modeling. One main algorithm for topic modeling is LDA, which is an unsupervised learning method that identifies the topics based on the weights of words within the document (Güven, 2020). At this step of the research, LDA algorithm was preferred. In the LDA method, the researcher can determine the total number of topics to be identified from the document. (Okcu et al., 2019). As an example, five topics that should preferably be determined from a document with negative election posts and 30 dominant words in these topics are shown below in Figure 5. After modeling, all the dominant topics in the posts are given in the findings section.

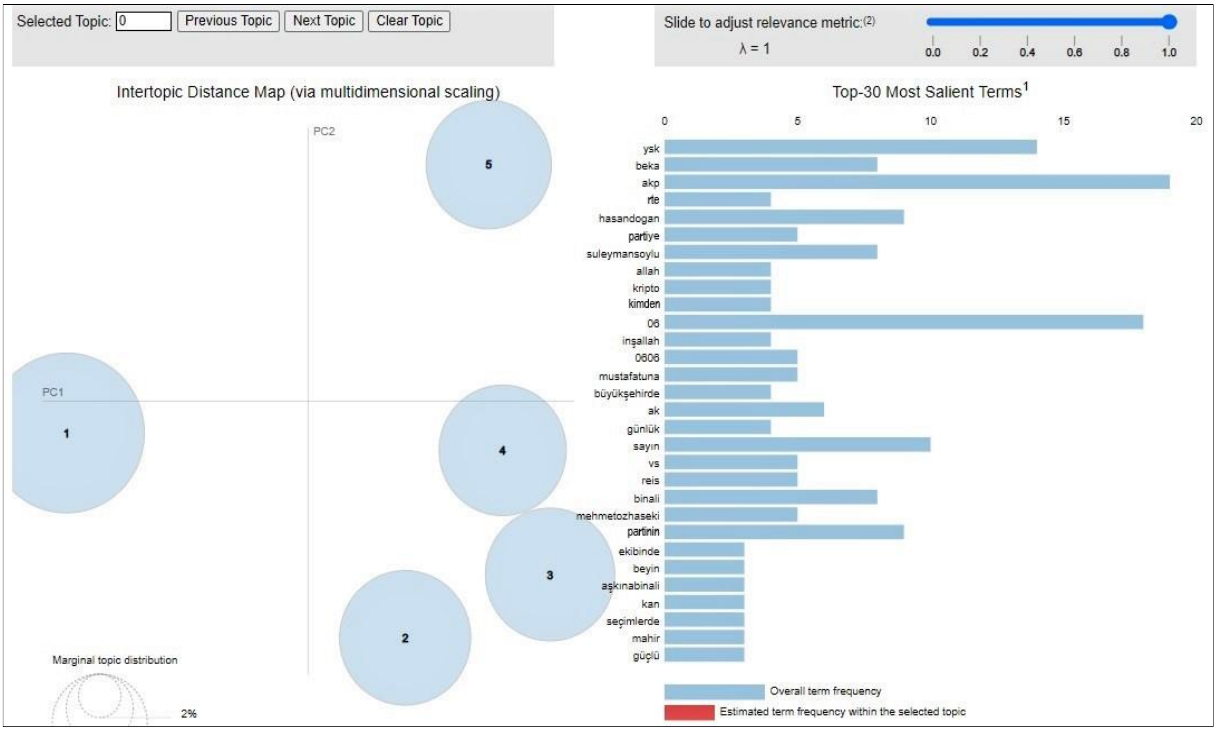

Figure 5. LDA model results example (five topics and terms featured in the topics)

2.7. Calculation of Candidate Representation/Popularity Ratio

There are various studies that deal with the representation/popularity ratio from the data obtained using X (Twitter) (O'Connor et al., 2010; Tumasjan et al., 2011; Soler et al., 2012; Ibrahim et al., 2015; Ramteke et al., 2016; Wang and Gan, 2018). To determine the representation/popularity ratios of the candidates in this study, the formula developed by Tumasjan et al. (2011) was used.

$$popularity = \frac{pos(a) + neg(b)}{pos(a) + neg(a) + pos(b) + neg(b)}$$

(a) the candidate whose popularity rate is to be calculated, *(b)* the other candidate, *pos:* the number of positive tweets and *neg:* the number of negative tweets.

3. Findings

3.1. Spatio-Temporal Post/Tweet Activities During Election Period

After the methods were used, a total of 65,635 posts covering the period of December 1, 2018 - April 20, 2019 (141 days) were collected on X (Twitter) (Table 2).

Table 2. Total number of posts collected by two methods (Twitter API and web scraping)

Twitter API	Web scraping			Total
	Point radius filter	Place id filter	Word pair query	
27,890	16,450	17,531	3,764	65,635

After data pre-processing and the keyword approach were applied to the 65,635 total posts, a total of 5,755 posts were identified. The sentiment analysis results of these posts and the dominant topics obtained after topic modeling will be examined in the next sub-title.

3.2. Understanding of Sentiment Analysis and Topic Modeling Results

A total of 2,269 posts for the People's Alliance were identified in the tweets we examined. The time-average sentiment graph created after the analysis applied to these posts is given below (Figure 6).

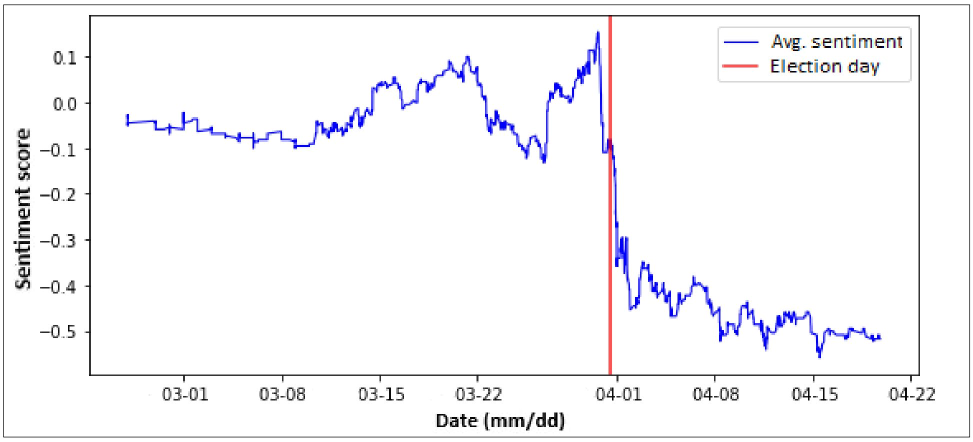

Figure 6. Sentiment state towards the People's Alliance

When the graph is analyzed, it can be seen that while the neutral posts held an important place for the alliance before the election day, the sentiment curve showed a negative trend with a significant increase in negative posts after the election day.

After topic modeling, among the positive posts about the People's Alliance, Recep Tayyip Erdoğan, the AK Party, and utilitarian demands (such as retirement age victims, shortened military service through payment, etc.) stand out as dominant topics. The topics in the negative posts vary, focusing on Recep Tayyip Erdoğan, the AK Party, Delegated Legislation, the Supreme Election Committee (SEC), and the Istanbul Metropolitan Mayor (IMM) Election (Figure 7).

POSITIVE	NEGATIVE
Recep Tayyip Erdoğan	Recep Tayyip Erdoğan
AK Party	AK Party
Wants in direction of benefit	Delegated Legislation Supreme
	Election Committee
	Istanbul MM Election

Figure 7. Topics included in posts towards the People's Alliance

For the other alliance (Nation Alliance), 1,147 posts were identified among the tweets. The time-average sentiment graph created after sentiment analysis is given below (Figure 8). This graph shows a distinct difference in daily sentiment state. The sentiment curve, which varied between neutral and negative until the election day, reached its highest positive level on the election day. However, after the election day, it reached its highest negative position. In general, negative sentiment was effective for this alliance.

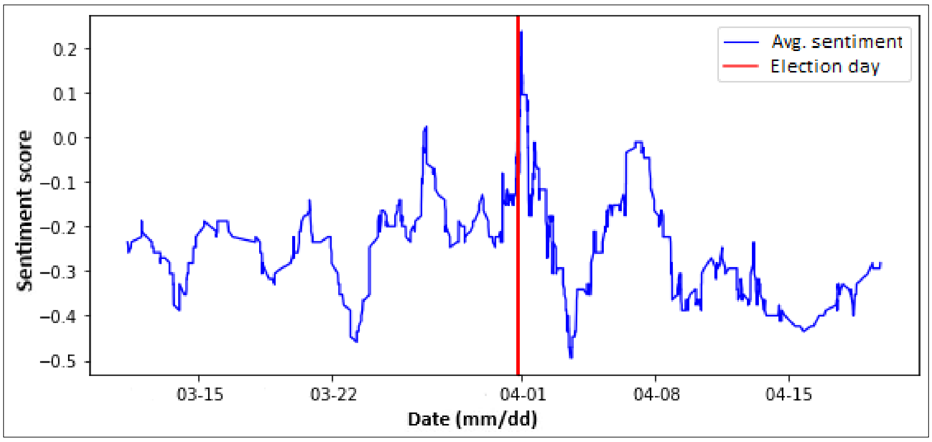

Figure 8. Sentiment state towards the Nation Alliance

Ekrem İmamoğlu, the IMM Election, and CHP members stand out in the positive posts about the Nation Alliance. In the negative posts about the alliance, Ekrem İmamoğlu, Mansur Yavaş, and the CHP-HDP relationship are particularly notable topics (Figure 9).

POSITIVE	NEGATIVE
Ekrem İmamoğlu	Ekrem İmamoğlu
İstanbul MM Election	Mansur Yavaş
CHP members	CHP-HDP relationship

Figure 9. Topics included in posts towards the Nation Alliance

There are 542 posts created for Mehmet Özhaseki, the candidate for the People's Alliance Ankara Metropolitan Mayor (AMM), and the previous AK Party Ankara Metropolitan Administration (AMA). The time-average sentiment graph formed by the related posts is shown below (Figure 10). When examining the graph, it is noticeable that mostly neutral posts were prevalent during the preelection days, while negative sentiment dominated the post-election days. Accordingly, the graph shows a significant similarity to the graph created for the People's Alliance.

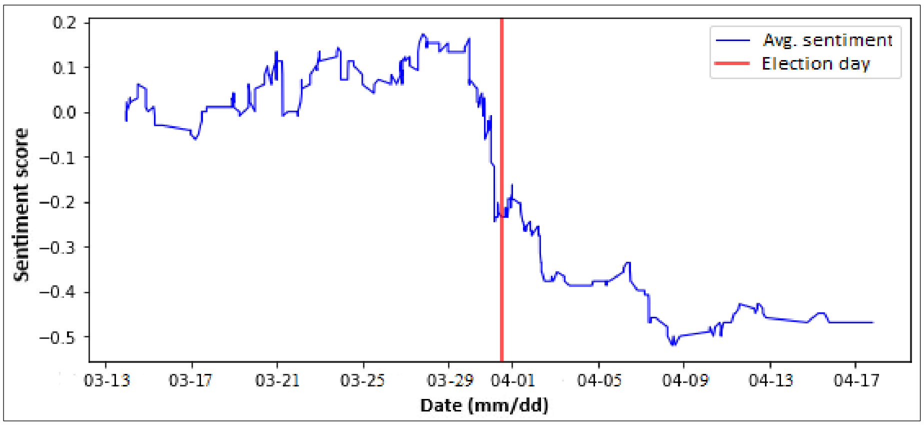

Figure 10. Sentiment state towards Mehmet Özhaseki and the previous AK Party AMA

After the analysis, it is evident that positive posts mainly revolve around candidate Mehmet Özhaseki and former mayor Melih Gökçek. On the other hand, negative posts include the previous mayor Melih Gökçek, activities of the previous AK Party AMA, and the 2014 Local Election (Figure 11).

POSITIVE	NEGATIVE
Mehmet Özhaseki	Melih Gökçek
Melih Gökçek	Activities of Ankara MM
	2014 Local Election

Figure 11. Topics included in posts towards Mehmet Özhaseki and the previous AK Party AMA

The time-average sentiment graph formed by the posts for the Nation Alliance AMM candidate Mansur Yavaş is displayed below (Figure 12). When examining the graph, it can be observed that the number of neutral posts was high before the election, but positive posts increased on the day of the election and continued to rise for one more week.

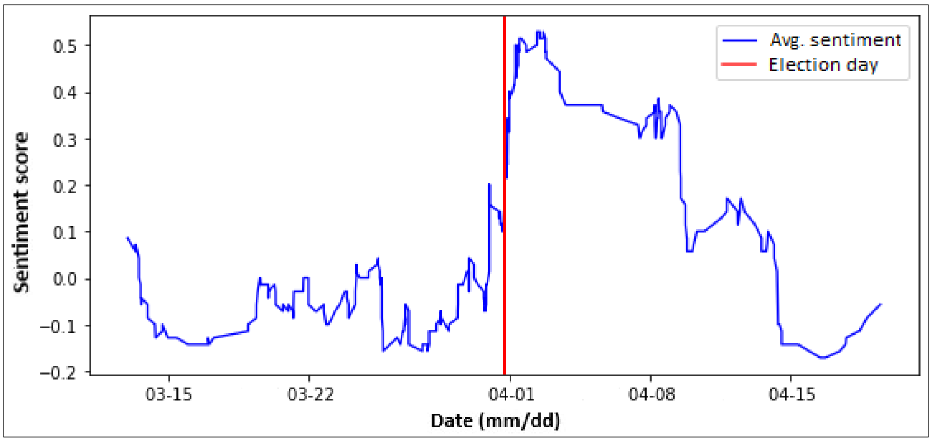

Figure 12. Sentiment state towards Mansur Yavaş

Positive posts for the Nation Alliance AMM candidate Mansur Yavaş include various hashtags with supportive and positive words for the candidate. Additionally, hashtags related to the 2014 Local Election are also present in the posts. In negative posts about Mansur Yavaş, it is evident that there are negative news articles about the candidate reported in the press and concerns raised about the CHP-HDP relationship (Figure 13).

POSITIVE	NEGATIVE
Supportive and positive words and hashtags	Negative news about the candidate
2014 Local Election	CHP-HDP relationship

Figure 13. Topics included in posts towards Mansur Yavaş

3.3. Popularity Ratios and Actual Votes of Candidates

According to the results of the March 31, 2019 Local Election, Mansur Yavaş was preferred as the Ankara Metropolitan Mayor with 57.71% of the votes in Yenimahalle. Mehmet Özhaseki was the candidate with the second highest number of votes for Metropolitan Mayor.

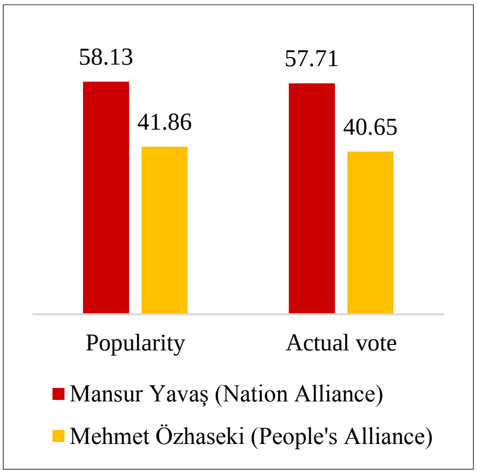

Figure 14. Popularity and actual vote rates of the candidates (%)

The popularity rates for the elected candidate and the candidate with the second highest votes are shown above in Figure 14. When we analyze the popularity of the candidates, we can observe a significant correlation between the popularity rates and the actual vote rates.

4. Discussion and Conclusion

It is observed that individuals create their own posts about elections and election-related elements on X (Twitter) for different purposes and reasons. These posts contain information about voting behavior and preferences, expressed through phrases such as *"I will,"* or *"I will not" vote*, *"I have voted,"* or *"have not voted,"* and *"I have preferred" or "have not preferred"* a particular administration or candidate. When examining the posts more closely, it has been determined that there are various factors that influence voting preferences, stemming from both socio-political and geographical contexts.

In posts created for alliances, the candidates' images, political parties, leaders, institutes, managers, country agenda, and utilitarian demands come to the fore. Candidates constitute the most important factor in the posts about the election of the metropolitan mayors. In addition, the previous mayor, metropolitan municipality services, party and agenda factors, and historical-political events in the space over which would-be elected officials are seeking to govern are effective in the creation of posts. To sum up, national and institutional topics are of great importance in election posts. This conclusion is similar to the view put forward by Agnew (1996) and supported by Axenov and Papadopoulos (1997), that national, institutional effects are reinterpreted (nested) by individuals in space. Also, we understand that voters reflect historical-political experiences/events in their election posts. This effect can be associated with Agnew's (1996) historical-geographical context. Based upon popularity calculation results, it has also been determined that there is a significant closeness between the popularity rates of the candidates on the social network and the actual vote rates. This finding has revealed a significant potential in predicting election results through X (Twitter).

Finally, as seen in the research results, thanks to its features, X (Twitter) provides an important opportunity for election communication and interpersonal information flow. This opportunity not only can bring together individuals with common or different views living in the same space in a virtual environment, but it also allows them to talk about various topics, factors, demands, and individuals. The data developed as individual voters harness this opportunity, and the methodology used to obtain and make sense of this data, are important for understanding voter behavior and spatial differences in voting, as well as for revealing contextual effects on voting.

These arguments can also be instructive for researchers in electoral geography who focus on Turkey's elections. At the time of writing this article, the general elections of 2023 were highly contested, and the People's Alliance emerged as the victor. Two major alliances competed in this election: the People's Alliance (AK Party and MHP) and an expanded Nation Alliance (CHP, IYI Party, Democrat Party, Democracy and Progress Party, Felicity Party, and Future Party), popularly known as the "table of six". The approach proposed in the study can reveal the spatial patterns of public attitudes and opinions towards alliances and presidential and mayoral candidates, as well as enable vote forecasting for candidates. In fact, since individuals who support an alliance may hold dissenting ideas about some of the parties and candidates who have come together within one or the other alliance, candidate and party performances can also be emphasized in future studies. Additionally, one can focus on the reasons behind attitudes, opinions, and voting preferences. This is because individuals' posts during the election period include many factors that arise from the local, national, and historical context.

In future research, examining the effect of interpersonal information flow in social networks and the potential of social networks to produce a micro-sociological place effect will make important contributions to the field of electoral geography. This can be linked to research in electoral geography, which addresses the interpersonal information flow and the micro-sociological place effect (Agnew, 1996; Cox, 1969; Johnston and Pattie, 2006; O'Loughlin, 1981). However, our approach - illustrated through the case of the 2019 Local Elections in Turkey - suggests the existence of new tools for studying these dynamics. Contexts such as workplaces, neighborhoods of residence, schools, etc., can clearly have an impact on voting and the political decisions made by individuals. Social networks, thanks to their features, have the ability to unite individuals with similar or diverse views, creating a space for mass and individual communication and interaction. Therefore, it is valuable to consider the potential inclusion of social networks in the aforementioned contexts.

Notes

1. Due to recent changes in X policies, the free X API products (Stream, Rest, and so on) have been discontinued. Instead, X (Twitter) API v2 products are now available in the Free, Basic, Pro, and Enterprise tiers. For more detail: https://developer.twitter.com/en/docs/twitter-api

Acknowledgement and Information

This article is derived from the first author's master's thesis entitled "*Electoral Geography with Social Media Mining: Ankara in the Example of the March 31, 2019 Local Election*" with the number 690263 CHE (YÖK).

We would like to extend our thanks to the editor and anonymous reviewers for their thoughtful suggestions to enhance the quality of the article. Additionally, we extend our thanks to Dr. Nicholas Crane from the University of Wyoming for his valuable contribution in improving this article.

References

Adams, P. C. (2015). Social media.In Agnew, J.A., Mamadouh, V., Secor, A. J., Sharp, J. (Eds.), *The Wiley Blackwell Companion to Political Geography*, West Sussex: John Wiley & Sons. 393-406.

Agnew, J. A. (1996). Mapping politics: How context counts in electoral geography. *Political Geography*, 15 (2), 129–146. doi: 10.1016/0962-6298(95)00076-3

Axenov, K. E., Papadopoulos, A. G. (1997). Long-term tendencies in the electoral behavior and the geography of voting in St. Petersburg: 1989–1995. *GeoJournal*, 42 (4), 433-448. doi: 10.1023/A:1006854916286

Bayraktar, S. U. (2011). Yerelin seçimi, çözümlemenin yereli: Mersin'de Mart 2009 il genel meclisi ve büyükşehir belediye başkanlığı seçimleri. *Toplum ve Bilim*, (121), 9-47.

Bekaroğlu, E. A., Kaya Osmanbaşoğlu, G. (2021). Introduction to Turkey's electoral geography: An overview since 1950. In Bekaroğlu, E. A., Kaya Osmanbaşoğlu, G. (Eds.), *Turkey's Electoral Geography Trends, Behaviors, and Identities*, Oxon-New York: Routledge. 1-23.

Belcastro, L., Cantini, R., Marozzo, F., Talia, D., Trunfio, P. (2020). Learning Political Polarization on Social Media Using Neural Networks. *IEEE Access*, 8, 47177-47187. doi: 10.1109/ACCESS.2020.2978950.

Bermingham, A., Smeaton, A. (2011). On using Twitter to monitor political sentiment and predict election results. In Bandyopadhyay, S., Okumura, M. (Eds.) *Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011)*, Chiang Mai: Asian Federation of Natural Language Processing, 2-10.

Books, J. W., Prysby, C. L. (1991). *Political Behavior and The Local Context.* New York: Praeger Publishers.

Cantey, C. (2013). *Data-mining social media for spatiotemporal patterns of negative opinion*. Master's thesis. Wisconsin: University of Wisconsin–Madison, College of Letters and Science.

Ceron, A., Curini, L., Iacus, S. M., Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 16(2), 340-358. doi: 10.1177/1461444813480466

Chauhan, P., Sharma, N., Sikka, G. (2023). Application of Twitter sentiment analysis in election prediction: a case study of 2019 Indian general election. *Social Network Analysis and Mining*, 13(1), 1-29. doi: 10.1007/s13278-023-01087-8

Convery, I., Cox, D. (2012). A review of research ethics in internet-based research. *Practitioner Research in Higher Education*, 6 (1), 50-57. Retrieved from http://insight.cumbria.ac.uk/id/eprint/1334/

Cox, K. R. (1969). The voting decision in a spatial context. *Progress in Geography* 1, 81-117.

Çarkoğlu, A., Avcı, G. (2002). An analysis of the electorate from a geographical perspective. In Sayari S., Esmer, Y. (Eds.), *Politics, parties, and elections in Turkey*, Boulder-London: Lynne Rienner Publishers. 115-136.

Çarkoğlu, A., Hinich, M. J. (2006). A spatial analysis of Turkish party preferences. *Electoral Studies*, 25, 369-392. doi: 10.1016/j.electstud.2005.06.010

GeeksforGeeks (2023). What is web scraping and how to use it? Access date: 28.12.2023, Retrieved from https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4), 211-221. doi: 10.1007/s10708-007-9111-y

Gundecha, P., Liu, H. (2012). Mining social media: a brief introduction. *Tutorials in Operations Research*, 1-17. doi: 10.1287/educ.1120.0105

Güven, A. (2020). Latent Dirichlet Allocation (LDA) Algoritması. Retrieved from https://medium.com/@anilguven1055/latent-dirichlet-allocation-ldaalgoritmas%C4%B1 13154d246e05

Han, J., Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Massachusetts: Morgan Kaufmann Publishers.

Harrison, T. M., Barthel, B. (2009). Wielding new media in Web 2.0: Exploring the history of engagement with the collaborative construction of media products. *New Media & Society*, 11 (1-2), 155-178. doi: 10.1177/1461444808099580

IBM. (2021). Data Mining. Retrieved from https://www.ibm.com/cloud/learn/data-mining

Ibrahim, M., Abdillah, O., Wicaksono, A. F., Adriani, M. (2015). Buzzer detection and sentiment analysis for predicting presidential election results in a Twitter nation. *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Los Alamitos, CA: IEEE Computer Society. 1348-1353.

İncioğlu, N. (2002). Local elections and electoral behavior. Sayari, S., Esmer, Y. (Eds.), *Politics, Parties, and Elections in Turkey*, Boulder-London: Lynne Rienner Publishers. 73-90.

Jaidka, K., Ahmed, S., Skoric, M., Hilbert, M. (2019). Predicting elections from social media: a three-country, three-method comparative study. *Asian Journal of Communication*, 29(3), 252-273. doi: 10.1080/01292986.2018.1453849

Johnston, R. J., Pattie, C. (2006). *Putting Voters in Their Place: Geography and Elections in Great Britain*. Oxford: Oxford University Press.

Kalaycıoğlu, E. (1999). The shaping of party preferences in Turkey: Coping with the Post-Cold War era. *New Perspectives on Turkey*, (20), 47-76. doi: 10.1017/S0896634600003137

Kamalak, İ. (2013). Yerelin yerel seçimlere etkisi: belediye başkan adayları üzerinden bir inceleme. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 15 (3), 419-446. Retrieved from https://dergipark.org.tr/en/pub/deusosbil/issue/4632/63145

Khan, A., Boudjellal, N., Zhang, H., Ahmad, A., Khan, M. (2023). From social media to ballot box: Leveraging location-aware sentiment analysis for election predictions. *Computers, Materials & Continua*, 77(3), 3037-3055. doi: 10.32604/cmc.2023.044403

Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39 (1), 61-72. doi: 10.1509/jmkr.39.1.61.18935

Köksal, A. (2020). *BERT-Sentiment-analysis-Turkish*, GitHub repository. Retrieved from https://github.com/akoksal/BERT-Sentiment-Analysis-Turkish

Liu, R., Yao, X., Guo, C., Wei, X. (2021). Can we forecast presidential election using Twitter data? An integrative modelling approach. *Annals of GIS*, 27(1), 43-56. doi: 10.1080/19475683.2020.1829704

Markham, A. N. (2004). Internet communication as a tool for qualitative research. In Silverman, D. (Ed.), *Qualitative research: Theory, method and practice* (2nd Ed.), London: Sage Publications. 95-124.

Marozzo, F., Bessi, A. (2018). Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining* 8 (1). doi: 10.1007/s13278-017-0479-5

McPhail, I. R. (1971). Recent trends in electoral geography. In Johnston, R. J., Chapman, J., Soons, J. M. (Eds.), *Proceedings of the Sixth New Zealand Geography Conference*, Christchurch: New Zealand Geographical Society, 7-12.

Miş, N., Duran, H. (2019). İttifaktan blok siyasetine 31 Mart yerel seçimleri. *Adam Akademi Sosyal Bilimler Dergisi*, 9 (2), 363-393. doi: 10.31679/adamakademi.634750

Murugesan, S. (2007). Understanding Web 2.0. *IT Professional*, 9 (4), 34-41. doi: 10.1109/MITP.2007.78

O'Connor, B., Balasubramanyan, R., Routledge, B., Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of The Fourth International AAAI Conference on Weblogs and Social Media*, California: The AAAI Press. 122-129.

Oğlakcı, B., Uzun, A. (2021). Metinsel mesajdan mekânsal bilgiye: coğrafya araştırmalarında veri kaynağı olarak sosyal ağ Twitter. *lnternational Journal of Geography and Geography Education*, (44), 176-191. doi: 10.32003/igge.927907

Oikonomou, L., Tjortjis, C. (2018). A Method for Predicting the Winner of the USA Presidential Elections using Data extracted from Twitter. *South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM)*, 1-8. Doi: 10.23919/SEEDA-CECNSM.2018.8544919.

Okcu, M., Albayrak, M., Topal, K., Turhan, G., Altıntaş. (2019). Yapay zekayla sosyal medyadan şehir verisini dinlemek. *Şehircilik ve Mutlu Şehirler Uluslararası Akademik Bildiriler Kitabı,* Kocaeli: Kocaeli BB. 302-321.

O'Loughlin, J. (1981). The neighbourhood effect in urban voting surfaces: a crossnational analysis. In Burnett, A. D., Taylor, P. J. (Eds.), *Political Studies from Spatial Perspectives: Anglo-American Essays on Political Geography*, Chichester: John Wiley. 357-388.

Olshannikova, E., Olsson, T., Huhtamäki, J., Kärkkäinen, H. (2017). Conceptualizing big social data. *Journal of Big Data*, 4(1), 1-19. doi: 10.1186/s40537-017-0063-x

Özen, I. C., Kalkan, K. O. (2017). Spatial analysis of contemporary Turkish elections: a comprehensive approach. *Turkish Studies*, 18, 358-377. doi: 10.1080/14683849.2016.1259576

Pattie, C., Johnston, R. J. (2009). Electoral geography. In Kitchin, R., Thrift, N. (Eds.), *International Encyclopedia of Human Geography*, Elsevier. 405–422.

Ramteke, J., Shah, S., Godhia, D., Shaikh, A. (2016). Election result prediction using Twitter sentiment analysis. 2016 *International Conference on Inventive Computation Technologies (ICICT)*, Piscataway, NJ: IEEE. 1-5.

Reynolds, D.R. (1990). Whither electoral geography? A critique. In Johnston, R. J., Shelley, F.M., Taylor, P.J. (Eds.), *Developments in Electoral Geography,* London: Routledge. 22-35.

Sakaki, T., Okazaki, M., Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. *Proceedings of the 19th International Conference on World Wide Web*, New York: Association for Computing Machinery. 1-10.

Scanfeld, D., Scanfeld, V., Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, 38 (3), 182-188. doi: 10.1016/j.ajic.2009.11.004

Seymour, W. S. (2001). In the flesh or online? Exploring qualitative research methodologies. *Qualitative Research*, 1 (2), 147-168. doi: 10.1177/146879410100100203

Shelley, F. M. (2006). Electoral geography. In Warf, B. (Ed.), *Encyclopedia of Human Geography*, California: SAGE Publications. 127-128.

Shin, M. (2015). Electoral geography in the twenty- first century. In Agnew, J. A., Mamadouh, V. Secor, A. J., Sharp, J. (Eds.) *The Wiley Blackwell Companion to Political Geography*, West Sussex: John Wiley & Sons. 279-296.

Snelson, C. L. (2016). Qualitative and Mixed Methods Social Media Research: A Review of the Literature. *International Journal of Qualitative Methods*, 15(1). doi: 10.1177/1609406915624574

Soler, J. M., Cuartero, F.i Roblizo, M. (2012). Twitter as a tool for predicting elections results. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Los Alamitos, CA: IEEE Computer Society. 1194-1200.

Song, M., Kim, M. C., Jeong, Y. K. (2014). Analyzing the political landscape of 2012 Korean Presidential Election in Twitter. *IEEE Intelligent Systems*, 29 (2), 18-26. Doi: 10.1109/MIS.2014.20.

Şahin, S. Z., Çekiç, A., Gözcü, A. (2014). An attempt of a local government monography method in Ankara: The example of Çankaya Municipality. *Journal of Ankara Studies*, 2 (2), 159-183.

Şahin, S. Z., Gözcü, A. (2017). Monograph of Yenimahalle Municipality. *Journal of Ankara Studies*, 5 (2), 283-318.

Temple, L. (2023). Reckoning with the digital turn in electoral geography. *Progress in Human Geography*, 47(4), 555-574. doi: 10.1177/03091325231170328

Tumasjan, A., Sprenger, T. O., Sandner, P. G., Welpe, I. M. (2011). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Social Science Computer Review*, 29 (4), 402-418. doi: 10.1609/icwsm.v4i1.14009

Tweet Location Metadata, (2023). Tweet with Twitter Place. Access date: 05.08.2023. Retrieved from https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/geo#:~:text=Tweet%20Location%20Metadata&text=There%20are%20two%20'root%2Dlevel,a%20Tweet%3A%20coordinates%20and%20place%20.&text=The%20place%20object%20is%20always,is%20assigned%20an%20

exact%20location.

Wang, L., Gan, J. Q. (2018). Prediction of the 2017 French election based on Twitter data analysis using term weighting. *10th Computer Science and Electronic Engineering (CEEC 2018),* Piscataway, NJ: IEEE. 231-235.

Wuthrich, F. M. (2015). *National Elections in Turkey: People, Politics, and The Party System*. Syracuse: Syracuse University Press.

Yıldırım, A., Şimşek, H. (2016). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. (10. Ed.). Ankara: Seçkin Yayıncılık.

Zafarani, R., Abbasi, M. A., Liu, H. (2014). *Social Media Mining: An Introduction*. New York: Cambridge University Press.