Location-based Sentiment Analyses and Visualization of Twitter Election Data

USSAMA YAQUB, NITESH SHARMA, and RACHIT PABREJA, Rutgers Business School
SOON AE CHUN, City University of New York
VIJAYALAKSHMI ATLURI and JAIDEEP VAIDYA, Rutgers Business School

In this article, we perform sentiment analyses of Twitter location data. We use two case studies: US presidential elections of 2016 and UK general elections of 2017. For US elections, we plot state-wise user sentiment towards Hillary Clinton and Donald Trump. For UK elections, we download two disparate datasets, using keywords and location coordinates, looking for similar tendencies in sentiment towards political candidates and parties. The overall objective of the two case studies is to evaluate similarity between sentiment of location-based tweets and on-ground public opinion reflected in election results. We discover Twitter location sentiment does indeed corroborate with the election result in both cases. We also discover similar tendencies in Twitter sentiment towards political candidates and parties regardless of the methodology adopted for data collection.

CCS Concepts: • **Information systems** → Information systems applications; Spatial-temporal systems; Location based services;

Additional Key Words and Phrases: Twitter, location data, sentiment analysis

1 INTRODUCTION

Social media platforms, such as Facebook and Twitter, have not only changed the way we interact with one another but also the way we share news and comment on world events. In June 2019, Twitter had more than 348M monthly active users creating around 500M tweets a day [36]. This high level of public connectedness has empowered users to start trends and make some news more popular than others, changing the way in which events are reported.

One area that has witnessed an increase in use of Twitter is that of electoral campaigning. With growing importance of Twitter as a communication tool, politicians and political parties now maintain active presence on the micro-blogging platform. This surge in Twitter usage for political campaigning during elections has also increased interest of researchers. Election prediction using Twitter data is becoming a popular research area [34, 38] where citizen sentiment is utilized to estimate candidate performance in general elections [43].

However, there are still supplementary Twitter meta-data that have not been fully used in data analyses. For example, while measures such as followers, re-tweets, and so on, have been looked at intensively in Twitter data analysis, utility features have mostly been overlooked [40]. One such feature is that of location. Twitter allows its users to provide location information in their profile. Although this attribute is optional, a significant number of users still choose to give their city/state and country information voluntarily. Users can also allow Twitter to access their device's location. In this case, Twitter utilizes the device's GPS for accurate latitude and longitude coordinates of user location. However, a very small number of users, usually less than 1%, enable this feature [10].

While a large number of election-related studies have utilized Twitter data to predict elections and evaluate candidate popularity, few have used location parameters [44]. We believe that through user location information, we can get a more accurate understanding of the on-ground public opinion. For example, recently in some elections around the world, a strong political divide is observed between urban and rural areas or between services-based and manufacturing-based zones where economic policies have unequal impact on the regions within a country. By utilizing location data, we can accurately gauge support levels for candidates and policies on each region or part of the country, giving us a more detailed picture of public opinion.

In this article, we perform sentiment analyses of Twitter data with respect to the location of tweets. Here, we utilize data downloaded during US elections of 2016 and UK general elections of 2017. We used Twitter streaming API to download the datasets for the two elections. We use location information, which is present as tweet meta-data, for sentiment analyses.

For US elections, we utilize user-provided location to tag tweets using our own parser and then map the aggregated sentiment of these tweets. We then create visualizations by plotting public sentiment towards the two presidential candidates, Hillary Clinton and Donald Trump, on a map of the US. Sentiment analysis is used to observe the attitudes, opinions, views, and emotions from text using Natural Language Processing (NLP). Thus, although sentiment analysis is primarily associated with classifying opinions in text into categories such as "positive," "negative," or "neutral," it also involves subjectivity analysis, where parts of speech such as adjectives, adverbs, and some group of verbs and nouns are taken as indicators of a subjective opinion [20]. Hence, through subjectivity analysis, we determine how opinionated a sentence is. However, polarity analyses of the text ascertains whether it expresses a positive or negative opinion, determining the emotional attitude of the text writer with respect to the topic under discussion [23]. In case of US elections 2016, we perform both subjectivity and polarity analyses of Twitter data.

For UK elections, we downloaded two disparate datasets, one using election-related keywords and the second using location coordinates of the tweets. First dataset is downloaded using keywords such as names of two major political parties *Labour* and *Tory*, their leaders *Jeremy Corbyn* and *Theresa May*, along with the term *GE2017*. The second dataset was downloaded by capturing tweets originating from within a bounding box created around Great Britain. In this case, tweets from only those users were captured who had enabled Twitter to access their device's location and were tweeting from within Great Britain. Tweets for both datasets were downloaded during the same days and hours. Here, we compare the sentiment regarding elections and candidates between location and keywords corpus. The purpose of this study is to generalize public sentiment towards candidates and political parties through data downloaded using different filters. Similar sentiment tendencies in the two datasets would indicate consistent public opinion on Twitter towards political leaders and parties regardless of the methodology employed to download data.

Thus, in this article, we make the following contributions:

- Evaluate utility of Twitter location data in terms of similarity with on-ground public opinion reflected in election results.
- Visualization and analyses of US elections 2016 location data, observing effectiveness of plotting sentiment of Twitter users on a map for visualization.
- Sentiment analyses of tweets downloaded using election-related keywords and location coordinates during UK elections of 2017, enabling us to generalize our findings by observing similar sentiment tendencies towards candidates and political parties regardless of the data download methodology.

In the next section, we present related work in the area of Twitter data analysis, while in Sections 3 and 4, we present the research questions for US and UK elections of 2016 and 2017, respectively. Section 5 discusses methodology and the Twitter datasets gathered during the two elections. In Sections 6 and 7, we present analyses of the two datasets. In Section 8, we discuss these results. In Section 9, we conclude our article.

2 RELATED WORK

Research activity in the area of Twitter sentiment analyses has grown steadily [9, 14, 42]. These analyses are applied to a variety of domains ranging from analyzing customer reviews to predicting events [2, 39]. One popular research area is that of gauging public sentiment towards government and elections. Here, research ranges from predicting resentment against government policies to forecasting election results [6, 9]. Increase in use of Twitter by politicians during general elections has resulted in increased research in evaluating the role of Twitter during electoral campaigns [1]. Similar research was also conducted in analyzing Twitter activity of US Congress members during their election campaigns where these candidates frequently posted information on Twitter regarding their political positions on various issues [16, 17].

Several measures have been proposed in research to predict comparative strength of political candidates on Twitter. One such measure is that of "relative support" used during the analysis of Spanish presidential elections of 2011 and Italian parliamentary elections of 2013 [6, 8]. The relative support measure utilizes time series slope of accumulated Twitter messages mentioning each candidate to compute their online support.

However, while there has been an exponential growth in the research pertaining to public opinion and sentiment prediction using Twitter, a debate still exists on the efficacy of the methods employed [15, 27]. As Twitter data do not contain important demographic information such as age, sex, income, education, and so on, doubts exist over how representative data sample is of general population.

Researchers have also examined the ways in which Twitter chatter can potentially influence mainstream media. It is shown that Twitter is now playing an important role as a news source for mainstream news [32]. Twitter is now used more than ever as a news agenda building tool for mainstream media [30, 31]. This was a very commonly observed phenomenon during the concluded US elections of 2016.

Social media has also enabled policy makers to collect data created by citizens for analysis [33]. Data deluge in the form of big data allows policy makers to develop a more sophisticated, wider-scale, finer-grained, and real-time understanding of their constituents [22]. Facebook, too, has become a popular tool with regards to communication between government institutions and citizens [19]. The majority of public sector organizations now manage Facebook pages through which they communicate with the community in which they operate. For example, almost all police departments of large US cities have presence on multiple social media platforms including Facebook and Twitter [7]. This presence enables them to quickly engage and get citizen help on any impending emergency. Local governments providing civic services such as garbage collection and infrastructure maintenance are also using social media to get feedback from citizens on priority of public works [25]. The idea of e-governance and smart city is taking hold globally and Internet has provided policy makers with the opportunity to increase government awareness and institutional responsiveness through citizen engagement. It

Table 1. Scale of Subjectivity and Polarity Scores

Metric	Explanation	Scale
Subjectivity	Measure of text as fact or opinion	0 to 1
Polarity	Text sentiment	−1 to +1

has been indicated that efforts by government to increase public participation in the affairs of government and taking greater input into decision making can improve public trust in government [21].

3 ANALYSIS OF US ELECTION TWEETS

In this section of the article, we develop research questions that are tested on Twitter data collected during the 2016 US presidential elections.

3.1 Subjectivity Analysis

Subjectivity analysis is a part of sentiment analyses where researchers classify a text in terms of the opinion expressed [20]. Presence of certain terms such as adjectives, adverbs, and some group of verbs and nouns are taken as indicators of a subjective opinion. Thus, subjectivity analysis is the classification of sentences as subjective opinions or objective facts. Hence, for a set of messages in a dataset, through subjectivity analysis, we identify sentences that are subjective from those that are objective. Speech patterns such as use of adjectives along with nouns are used as an indicator for subjectivity of a statement [20].

We used Python TextBlob for Natural Language Processing (NLP) to perform subjectivity analysis on our dataset. Python TextBlob is a popular NLP tool available for free that can be used to perform text subjectivity analysis. Subjectivity of every tweet is calculated with the help of Textblob library, which uses a built-in model to calculate subjectivity score. Table 1 shows the range of subjectivity values. A value close to 0 indicates an objective text while a value close to 1 indicates a highly subjective text. Working details of Python NLP are provided in Bird, Klein, and Loper 2009 [5].

The first research question relates with the analysis of Twitter location data to evaluate the subjectivity of user messages during the elections. There have been several studies on user behavior on Twitter during elections. Some of the research studies have suggested users are less than factual with their online comments and that they use Twitter as an echo chamber, where few opinions are restated repeatedly [12]. High level of recycling tweets is also a well-documented phenomenon. Many recent studies into Twitter analysis during general elections have observed that little original content was created by users during discussions and most engage in re-tweeting [6, 42]. Similarly, a high level of heterogeneous user behavior was also discovered on Twitter during the Venezuelan protests of 2010 [29].

Our purpose of analyzing user subjectivity is to evaluate whether Twitter users were tweeting or re-tweeting facts during political debate or whether most of their messages were emotional subjective opinions. Furthermore, Twitter content creation remains an interesting area of study as researchers analyze why some tweets become popular and are re-tweeted more than other messages [45]. This phenomenon is also detectable in political discussions where researchers have observed a high rate of content reuse [26].

Based on the above discussion, we wanted to analyze user content in terms of subjectivity. Hence, we propose the following research question:

Research Question 1: *How are subjectivity scores for each candidate varied across states and who was mentioned in more subjective tweets?*

To answer this question, we performed the following data analysis:

- Subjectivity analysis of the Twitter dataset using location attribute. Average subjectivity score for each candidate is calculated.

- Comparing the state-wise subjectivity scores for both candidates to gauge which contestant was mentioned in more subjective tweets.

3.2 Polarity Analysis

Polarity analysis is used to determine the emotional attitude of a text writer with respect to the topic under discussion [23]. With Twitter gaining popularity as a communication tool, interest in analyzing its data to determine public mood or opinion towards a topic or a popular personality has also grown [31, 38]. Examples of research studies utilizing Twitter sentiment range from election prediction to stock price forecast [28, 42]. Similarly, in terms of approval ratings, studies have discovered a strong correlation between Twitter sentiment and public opinion polls with some claiming as high as 80% correlation between the two [31].

Polarity analysis enables us to gain that insight by quantifying the sentiment of text. The text can thus be classified as negative, positive, or neutral. Several tools are available for sentiment analysis of short text; we have utilized Python TextBlob package for the polarity analysis on our dataset. Here, words in the lexicon are assigned scores for negative and positive polarity. These polarity scores range from -1 to 1, where -1 represents extremely negative sentiment while 1 represents extremely positive sentiment, respectively. A polarity score of 0 suggests a neutral sentiment. Table 1 briefly describes the polarity metrics. Using Python Textblob, polarity scores were assigned to tweets for both candidates. Textblob uses a built-in model to calculate the polarity values. These tweets were organized based on the states from which they originated.

We believe that polarity analysis taking location into consideration can help during elections by allowing candidates to focus on their areas of strength and weakness by allocating resources more efficiently and according to the geographical needs. Furthermore, metadata associated with each tweet are rich in detail and amongst other variables also include user location. In our study, we use this location attribute to evaluate the sentiment of both presidential candidates across the 10 most populous states of US. We then compare these sentiment scores of the two candidates with the election results. This will allow us to evaluate how indicative Twitter location data are of public opinion in these states. Hence, we have the following research question:

Research Question 2: *How does sentiment of Twitter messages based on state location correlate with the real-world sentiment of the public towards the two candidates?*

To answer this research question, we performed the following data analytical steps:

- Polarity analysis of the Twitter dataset organized into US states using location attribute. Average sentiment for both candidates is calculated for each state.
- Comparing this state-wise sentiment with the actual election result to test the accuracy of Twitter sentiment analysis based on location.

3.3 Subjectivity and Polarity Example

We will utilize a tweet by Donald Trump, posted during the election campaign, as an example to explain subjectivity, i.e., opinion and polarity:

- **Message:** @realDonaldTrump: So nice—great Americans outside Trump Tower right now. Thank you!
- **Subjectivity:** nice, great.
- **Polarity:** positive.

The above tweet ranks high in subjectivity and expresses a highly positive sentiment.

3.4 Case Selection

As it is not mandatory for Twitter users to share their location information, many tweets contain null or some arbitrary value for location variable. This greatly reduced the number of tweets used in our data analysis. In total, 3,341,324 tweets with valid user location values were gathered from all 50 states of the US for this study.

Table 2. Number of Tweets Collected from 10
Most Populous States of the US, Utilized for
Subjectivity and Polarity Analyses

State	Tweets	% of all tweets
California	471,501	14.11%
Florida	312,364	9.35%
Georgia	103,426	3.10%
Illinois	106,893	3.20%
Michigan	80,994	2.42%
New York	316,114	9.46%
North Carolina	103,421	3.10%
Ohio	98,593	2.95%
Pennsylvania	106,849	3.20%
Texas	314,021	9.40%
Total	2,014,176	60.28%

Although we use tweets from all states to construct US sentiment maps for both candidates shown in Figure 4 and 5, for our two research questions, we utilized tweets from only the 10 most populous states of the US. These 10 states contain almost 54% of the US population and accounted for over 60% of all tweets in our dataset. With a large population and thus a large Twitter user base, we were able to consistently gather a sizeable number of daily tweets from each of these states for both candidates. This allowed us to better determine average sentiment values for users tweeting from these states, as a small number of tweets could result in error due to an atypical sample. Table 2 shows the number of tweets collected from these 10 states of the US, which were used for state-wise subjectivity and polarity analysis.

We can observe an instance of this anomaly in Figure 4 where along with democratic strongholds of California, New York, Washington, Illinois, and so on, Hillary Clinton enjoys a high positive sentiment in South Carolina as well, which is a strong Republican state. This aberration is can be explained by the fact that very few location tweets were collected from South Carolina and further still, very few of those tweets mentioned Hillary Clinton. Hence, this sentiment is not an accurate representation of the ground reality. In the future, to cater for this inconsistency, we will gather data for a longer period of time, ensuring that a sufficiently large, representative sample is collected from all states.

4 ANALYSES OF UK ELECTIONS TWEETS

In this section, we evaluate the sentiment of the UK general elections of June 2017. Using the Twitter streaming API, we download data for 14 days during the general elections starting from May 27, 2017, to June 9, 2017. However, in this case, we downloaded two disparate datasets: one using election-related keywords and the other using location coordinates of tweets. The first dataset is downloaded using names of two major political parties and their prime minister candidates along with the term *GE2017*, which was a popular trend during the elections. The second dataset is downloaded by capturing tweets originating from within a bounding box created around Great Britain. Both these datasets were downloaded at the same time. We then compare the sentiment of location corpus with the keywords corpus as a whole, and for each election-related keyword individually.

In this case, we test the veracity of Twitter analysis to gauge public sentiment during elections. While as we see a rise in use of Twitter data analyses to gauge public opinion and, in some cases, even predict elections, many researchers have questioned the use of Twitter data for election prediction [15]. One criticism in this regard is that results of Twitter election prediction studies cannot be replicated [27]. Getting similar sentiment results from two different datasets collected during the same time period for the same event would suggest an overall

consistency. This allows us to generalize our findings regardless of the methodology used to download data, demonstrating the efficacy of Twitter data analyses to predict public sentiment towards candidates and political parties during general elections.

Another debate on using Twitter data analysis to predict elections is that of the filters used to download tweets. Twitter streaming API does not provide all tweets fulfilling the selection criteria. It only provides a subset, ranging anywhere from 1% to 40% of all tweets that fulfill the selection criteria. It is argued that choosing the terms of download can have a huge impact on the analysis results [30]. Hence, sentiment analyses of tweets containing similar terms in the location and keywords dataset will help us evaluate the impact of filters on sentiment analyses of election.

Thus, we propose the following research question:

Research Question 3: *How do sentiment of Twitter messages filtered on location and keywords correlate with one another, and do they indicate similar sentiment towards the candidates and political parties?*

To answer this research question, we performed the following data analytical steps:

- Sentiment analysis of two Twitter datasets downloaded using location and keyword filters during UK elections.
- Comparing sentiment of candidates and political parties in the two datasets, testing for significant similarity of differences.

5   METHODOLOGY

Our data-gathering and preparation methodology is composed of the following steps:

(1) Use keywords and location coordinates to gather tweets during US and UK elections of 2016 and 2017, respectively, using Twitter streaming API.
(2) Data cleaning and extraction.
(3) For the US elections, extraction of user location from tweet meta-data along with assigning polarity and subjectivity scores to tweets through NLP.
(4) For the UK elections, assigning sentiment scores to tweets downloaded using keywords and location co-ordinates utilizing SentiStrength.
(5) Formulation of research questions that are answered through data analyses and visualization.

Python is used to gather, clean, and extract user location from tweets. Python TextBlob is used for NLP to assign polarity and subjectivity scores to user messages for the tweets gathered during US elections, while SentiStrength software [35] is used to assign sentiment scores to the tweets gathered during UK elections. Finally, Tableau is used for mapping of candidate polarity on the map of the US. In the rest of this section, we will briefly discuss these methodology steps.

5.1   Datasets

We will now describe datasets gathered during US and UK elections. We will explain the criterion used to filter tweets along with the data cleaning and processing operations performed to prepare the data for analyses.

5.1.1   US Elections Dataset. US presidential elections were held on November 8, 2016. Here, we downloaded tweets for eight days starting from Nov 1st until election day (November 8th). These messages were downloaded using the keywords "Trump" and "Clinton," as they were the two primary presidential candidates. The location information of the users was extracted from the tweets, which are then categorized according to user state.

Although we downloaded over 15M user tweets from November 1st to November 8th, we were able to correctly extract 3,341,324 tweets using our location dictionary. However, while all 50 states were represented in our dataset, we decided to utilize the 10 largest states by population for our analyses. We restrict our analyses to

Fig. 1. Bounded rectangle was created around Great Britain using latitude and longitude coordinates. Tweets originating from within this area were captured in the location dataset.

Table 3. Number of Tweets Collected during UK Elections Using Election-related Keywords and Location Coordinates

UK Elections Datasets	Number of Tweets	Days
Downloaded using keywords	9,549,562	27th May—9th June
Downloaded using coordinates	2,770,051	27th May—9th June

these states, as they were well represented in our data and accounted for over 60% of all tweets in our dataset. Table 2 provides details on the number of tweets captured from these states.

5.1.2 UK Elections Dataset. Voting for UK general elections was held on June 8, 2017. Here, we downloaded data for 14 days starting from May 27, 2017 to June 9, 2017. As discussed earlier, for UK elections, we downloaded tweets using two different filter criteria. The first dataset was downloaded using keywords. In this case, tweets containing keywords such as "Jeremy Corbyn," "Theresa May," "Tories," "Labour," and "GE2017" were downloaded. We used names of the two major political parties and their heads along with the neutral terms of "GE2017" for data collection. In total, we downloaded around 9.5M tweets using terms.

The second dataset was downloaded using location coordinates of tweets. This was done by capturing tweets originating from within a bounding box created around Great Britain. This allowed us to capture tweets from only those users who had allowed Twitter to access their device's location and were tweeting within the latitude and longitude coordinates from Great Britain. Figure 1 shows the area from which tweets were downloaded. Around 2.7M tweets were downloaded using location coordinates. Although tweets for both terms and geo-tagged filters were downloaded for the same number of days and hours, number of tweets captured using location filters was considerably smaller than the keywords tweets (2.7M vs. 9.5M). This is due to the fact that a very small percentage of users allow Twitter to access their location [30]. Table 3 provides details of the two datasets.

Tweets for both datasets were downloaded during the same days and hours. Here, we compare public sentiment towards elections, political parties, and candidates, between the location and keywords corpus.
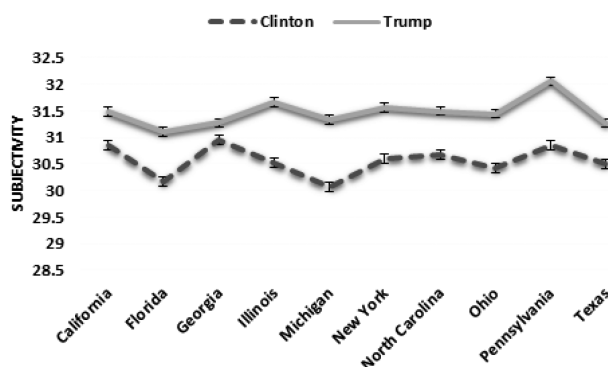
Fig. 2. Subjectivity scores of Clinton and Trump across 10 most populous states of the US.

For sentiment analyses, we utilize SentiStrength software [35] to calculate sentiment of each message. SentiStrength is a freely available tool that has been used to perform sentiment analyses in numerous research studies using Twitter data [9, 14]. The tool has been specifically developed to capture sentiment of short, informal texts [37].

5.2 Data Cleaning

While metadata of each tweet contain a multitude of attributes, we were interested in extracting only few of these variables such as tweet text, location, and time of tweet created. In case user state location was missing, we also created a data dictionary to map major cities such as New York, Los Angeles, Chicago, and so on, in their respective states.

Furthermore, presence of spam on Twitter in the form of automated bot activity is very well documented. Studies have indicated that almost 10.5% of all Twitter accounts might be automated, while as high as 9% of all tweets may be generated by these automated accounts [11, 18]. These bots are developed and designed to mimic human behavior and hence are hard to identify and remove from Twitter. Thus, along with missing fields, we also removed high activity accounts from our dataset.

6 ANALYSES OF US PRESIDENTIAL ELECTIONS 2016

In this section of the article, we present the results of our data analyses for US elections of 2016. We will utilize our findings to corroborate the research questions we developed in the previous section.

6.1 Subjectivity Analysis

We performed subjectivity analysis of tweets for both candidates using the Twitter location data. In subjectivity analysis, we determine how emotion, speculation, opinion, and sentiment are expressed in natural language [41]. While message polarity determines the positive or negative connotation of a text, subjectivity analysis tries to discern whether the text is subjective in the form of an opinion, belief, emotion, or speculation or objective as a fact [24].

We evaluate tweets from 10 states of the US during the 2016 elections in terms of their subjectivity. We wanted to gauge which of the two candidates had a higher subjectivity in the tweets mentioning them. Figure 2 displays the subjectivity scores of the tweets mentioning both candidates in the 10 most populous states of the US.

We can observe from Figure 2 that tweets mentioning Donald Trump have a higher subjectivity score for all 10 states compared with those mentioning Hillary Clinton. To further distinguish the divergence between subjectivity scores of the two candidates, we perform t-test on their subjectivity values. Following are our null and alternate hypotheses:
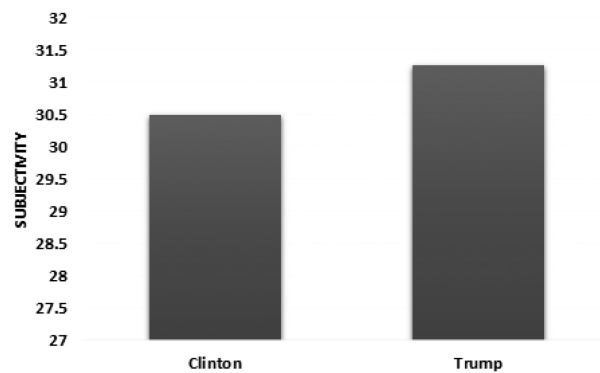
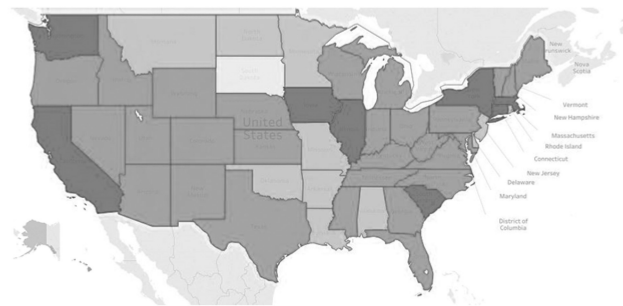Fig. 3. Overall subjectivity scores of Clinton and Trump.

Fig. 4. Sentiment of tweets mentioning Hillary Clinton across all states of the US. Darker shades indicate higher positive sentiment.

$H_0 -> Clinton_{Subjectivity} = Trump_{Subjectivity}$,
$H_1 -> Clinton_{Subjectivity} \neq Trump_{Subjectivity}$.

We used F test for sample variance and t-test for hypothesis testing. Test results are shown below:

t-stat	7.35
P(T<=t) two tail	7.96629E-07
t critical two tail	2.10

As the t-stat value is higher than t critical, we can reject the null hypothesis and accept the alternate hypothesis stating that tweets mentioning Donald Trump were more subjective than those mentioning Hillary Clinton. This phenomenon of high subjectivity of tweets mentioning Donald Trump was observed for all states of the US. Figure 3 shows overall subjectivity for both candidates across all states.

6.2 Polarity Analysis

Along with subjectivity analysis, we have also used location data to perform polarity analysis of tweets mentioning Hillary Clinton and Donald Trump. After all tweets are tagged with polarity scores, we calculate average polarity of each state using location of tweets. Figures 4 and 5 show the polarity map of US for Hillary Clinton and Donald Trump, respectively. Dark color indicates higher polarity and hence more positive sentiment for the candidate.
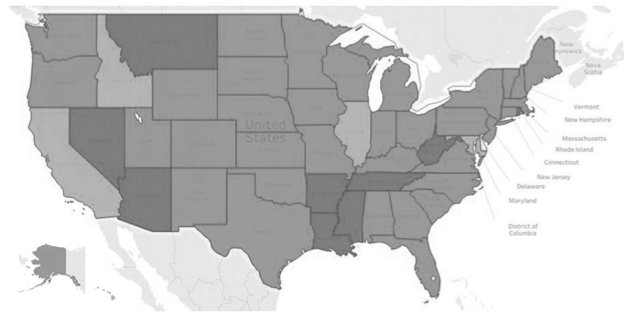
Fig. 5. Sentiment of tweets mentioning Donald Trump across all states of the US. Darker shades indicate higher positive sentiment.
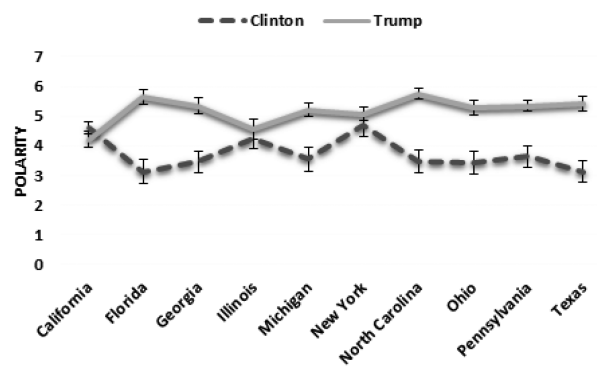

Fig. 6. Polarity of Clinton and Trump in 10 most populous states of the US. Higher polarity score indicates more positive sentiment.

For polarity comparison of the two candidates, just like subjectivity analysis, we utilize data from the 10 most populous states of the US, shown in Table 2. Through this analysis, we intend to identify sentiment of each state towards these candidates on Twitter. We want to evaluate how this polarity score compares with actual election results. Out of 10 states, Donald Trump won 7, while Hillary Clinton was victorious in 3. Figure 6 shows the polarity scores.

We observe Hillary Clinton as having higher polarity than Donald Trump only in the state of California. However, the states of Illinois and New York are within the margin of error for both candidates. Hillary Clinton won both of these states. Table 4 shows the polarity scores and error terms along with the eventual winner of these states in the elections.

It has been reported that Donald Trump enjoyed better sentiment on social media than Hillary Clinton [43]. He had more followers than Hillary Clinton and used Twitter more frequently with his tweets generating greater media attention. We can observe similar results in our location analysis where he enjoys better polarity score in 9 out of 10 states.

7 ANALYSES OF UK GENERAL ELECTIONS 2017

For UK elections, we perform comparative analyses of datasets downloaded using election-related keywords and location coordinates to evaluate similarities and differences in sentiment. As discussed earlier, through these analyses, we want to discern whether we can get similar sentiment results from two different Twitter datasets, collected during the same time period for the same event.

Table 4. Polarity Score and Standard Error of Tweets Mentioning Hillary Clinton and Donald Trump in 10 Most Populous States of the US

State	Clinton Polarity	Trump Polarity	Error Clinton	Error Trump	Election Winner
California	4.59	4.19	0.20	0.27	Clinton
Florida	3.12	5.65	0.40	0.25	Trump
Georgia	3.45	5.34	0.36	0.28	Trump
Illinois	4.26	4.55	0.35	0.32	Clinton
Michigan	3.55	5.20	0.41	0.23	Trump
New York	4.71	5.07	0.40	0.23	Clinton
North Carolina	3.46	5.74	0.37	0.19	Trump
Ohio	3.44	5.27	0.38	0.25	Trump
Pennsylvania	3.64	5.34	0.37	0.19	Trump
Texas	3.13	5.42	0.36	0.24	Trump

Higher polarity indicates higher positive sentiment. Last column of the table shows eventual winner of the state in 2016 presidential elections.
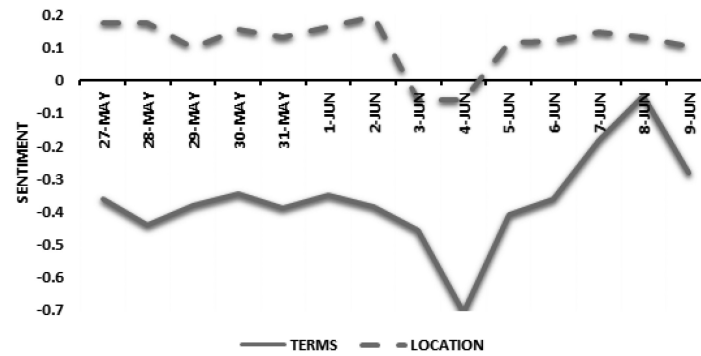

Fig. 7. Sentiment comparison of the two datasets collected during the UK general elections of 2017. Tweets in location dataset have more positive sentiment than tweets collected using keywords.

When we evaluate overall sentiment of the two datasets, we observe sentiment of location dataset to be consistently better than that of the keyword dataset. Figure 7 shows daily average sentiment of Twitter datasets collected using location coordinates of the UK and terms related with UK elections.

To further distinguish the divergence between sentiment of the datasets, we perform *t-test*. Following are our null and alternate hypotheses:

$H_0 -> Sentiment_{Terms} = Sentiment_{Location}$,
$H_1 -> Sentiment_{Terms} \neq Sentiment_{Location}$.

We used F test for sample variance and discover an F value of 3.46, which is greater than F critical value of 2.58. Hence, we apply *t-test* for two samples assuming unequal variance. Test results are shown below:

t-stat	11.58
P(T<=t) two tail	1.39232E-10
t critical two tail	2.08

As the t-stat value is higher than t critical, we reject the null hypothesis and accept the alternate hypothesis stating that sentiment of location dataset is significantly more positive than sentiment of keywords dataset.
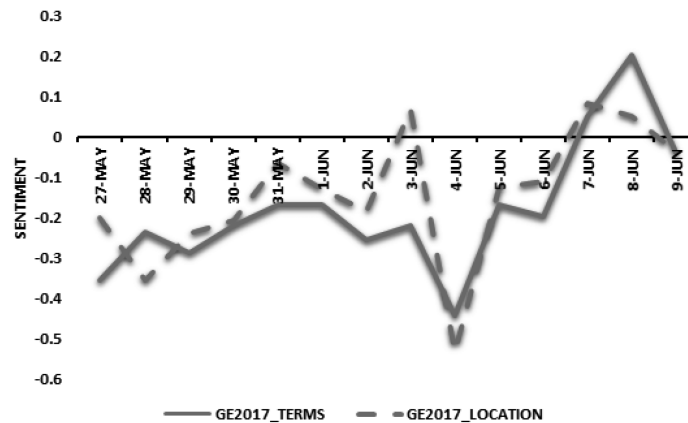
Fig. 8. Sentiment of keyword *GE2017* in the two datasets. Unlike Figure 7, sentiment of tweets containing GE2017 is very similar in both datasets.

However, for election-related sentiment comparison, we select tweets in the two datasets containing the keyword "GE2017." The term GE2017 is a reference to "general elections 2017" and was a popular trend during the UK elections. We used the keyword GE2017 to search for tweets in both datasets. As both datasets were downloaded during the same time period, we also had to eliminate tweets that were appearing in both datasets to make sure we have mutually exclusive sets. We discovered 1,288,656 tweets in the terms dataset and 43,699 tweets in the location dataset that contained the term GE2017. Figure 8 shows the average daily sentiment for tweets containing GE2017 in both datasets.

We can observe that messages containing GE2017, the sentiment of the two datasets is somewhat similar when compared with the sentiments of the overall datasets shown in figure 7.

Again, using *t-test*, we have the following null and alternate hypotheses:

$H_0 -> Sentiment\_GE2017_{Terms} = Sentiment\_GE2017_{Location}$
$H_1 -> Sentiment\_GE2017_{Terms} \neq Sentiment\_GE2017_{Location}$

We get an F value of 1.06, less than F critical value of 2.58. Hence, we apply *t-test* for two samples assuming equal variance. Test results are shown below:

t-test dataset sentiment.		*t-test* GE2017 sentiment.	
t-stat	10.79	t-stat	0.59
P(T<=t) two tail	8.6194 E-10	P(T<=t) two tail	0.56
t critical two tail	2.055	t critical two tail	2.055

In this case, the t-stat value is less than t critical and we cannot reject null hypothesis. The *t-test* failed to find evidence of difference between the two samples.

We perform similar sentiment analysis on both datasets. Figures 9 and 10 show sentiment of messages mentioning the heads of the two major political parties in the UK, Jeremy Corbyn and Theresa May. In the terms dataset, we discovered 2,738,337 tweets mentioning Jeremy Corbyn and 1,742,181 tweets mentioning Theresa May. In the location dataset, we found 25,231 tweets mentioning Jeremy Corbyn and 20,890 mentioning Theresa May. Table 5 shows the details of tweets for two candidates.

Figure 9 shows the daily average sentiment of tweets for them in the terms database, while Figure 10 displays the sentiment in the location dataset. We can see that in both datasets, which contain mutually exclusive
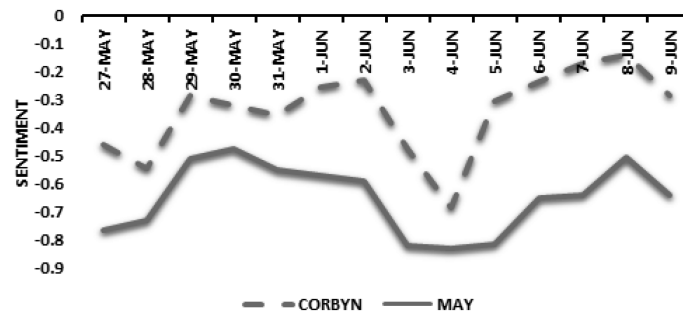
Fig. 9. Sentiment of Jeremy Corbyn and Theresa May in the Twitter dataset downloaded using keywords.
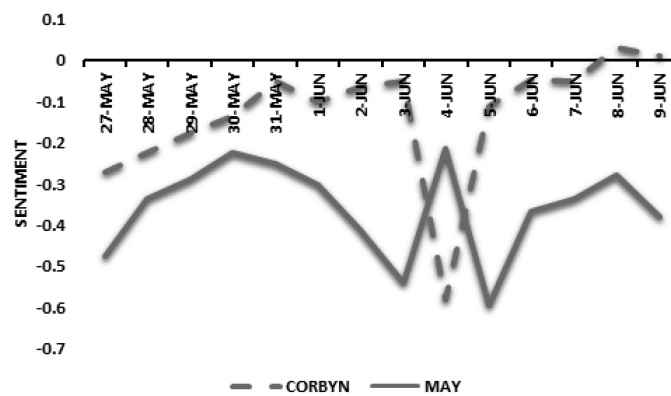

Fig. 10. Sentiment for Jeremy Corbyn and Theresa May in the Twitter dataset downloaded using location coordinates.

Table 5. Tweets Mentioning Heads of the Two Main Political Parties

Candidate	Tweets Keyword dataset	Sentiment	Tweets Location dataset	Sentiment
Jeremy Corbyn	2,738,337	− 0.297	25,231	− 0.058
Theresa May	2,043,767	− 0.569	20,890	− 0.34

Columns 3 and 5 provide the sentiment of tweets for keywords and location datasets, respectively. Jeremy Corbyn is mentioned in more tweets and has a more favorable sentiment than Theresa May in both datasets.

messages, Jeremy Corbyn enjoys a more favorable sentiment than Theresa May. We can further verify this claim by performing *t-test* on both sentiment values. We have the following null and alternate hypotheses:

$H_0 -> Corbyn_{sentiment} = May_{sentiment},$
$H_1 -> corbyn_{sentiment} \neq May_{sentiment}.$

We get an F value of 1.50 and 1.82 for terms and location datasets, respectively, both of which are less than F critical value of 2.58. Hence, we apply *t-test* for two samples assuming equal variance. Test results are shown below:

t-test terms dataset.	
t-stat	5.939
P(T<=t) two tail	2.88 E-06
t critical two tail	2.055

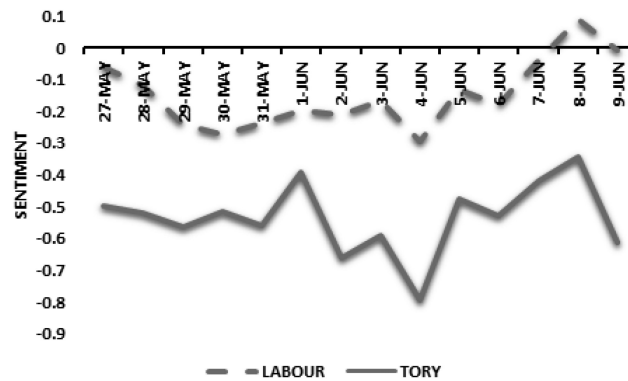| *t-test* location dataset.	
t-stat	4.397
P(T<=t) two tail	1.65 E-04
t critical two tail	2.055

Fig. 11. Sentiment of Labour and Conservative party in the Twitter dataset downloaded using keywords.
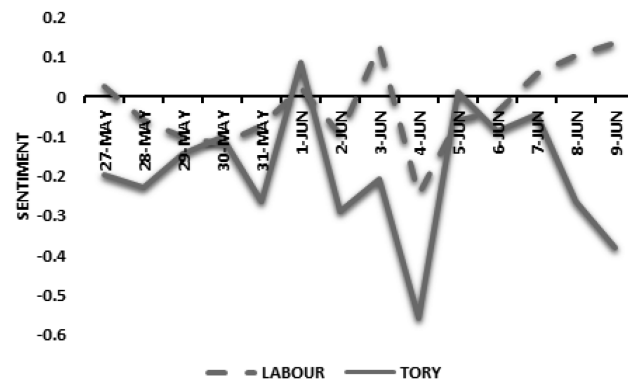

Fig. 12. Sentiment of Labour and Conservative party in the Twitter dataset downloaded using location coordinates.

Table 6. Tweets Mentioning the Two Main Political Parties, Labour and Tory

Party	Tweets Keyword dataset	Sentiment	Tweets Location dataset	Sentiment
Labour	2,351,996	-0.125	37,409	0.072
Tory	1,232,015	-0.514	13,354	-0.274

Columns 3 and 5 provide the sentiment of tweets for keywords and location datasets, respectively. We can observe that Labour is mentioned in more tweets and has more favorable sentiment than Tory in both datasets.

The results of our *t-test* verify that Jeremy Corbyn has a statistically significant better sentiment than Theresa May in both the datasets collected during the UK elections of 2017. Hence, we can reject the null hypothesis and can state that Corbyn enjoyed a better sentiment on Twitter during the elections when compared with May.

Along with the candidates, we searched both datasets for names of the two main political parties, Labour and Tory. Figures 11 and 12 show sentiment of messages mentioning Labour and Tory. In the terms dataset, we discovered 2,351,996 tweets mentioning Labour and 1,232,015 tweets mentioning Tory. In the location dataset, we found 37,409 tweets mentioning Labour and 13,354 mentioning Tory. Table 6 shows the details of tweets for two parties.

Figure 11 shows the daily average sentiment of tweets for the two parties in the terms dataset while Figure 12 displays the sentiment in location dataset. We can see here that in both datasets, Labour party enjoys a more favorable sentiment than Tory. We can further verify this claim by performing *t-test* on sentiment values of both parties. We have the following null and alternate hypotheses:

$H_0 -> Labour_{Sentiment} = Tory_{Sentiment},$
$H_1 -> Labour_{Sentiment} \neq Tory_{Sentiment}.$

We perform *t-test* paired two samples for mean. Test results are shown below:

t-test terms dataset.	
t-stat	14
P(T<=t) two tail	3.21E-09
t critical two tail	2.055

t-test location dataset.	
t-stat	3.643
P(T<=t) two tail	0.003
t critical two tail	2.16

The result of our *t-test* verifies Labour party as having significantly better sentiment than Tory in both datasets. Hence, we can reject the null hypothesis and can state that Labour party enjoyed a better sentiment on Twitter during the elections when compared with Tory.

It would be pertinent to mention here that Labour party did better than expected in the UK elections of 2017 [3]. Theresa May, prime minister and leader of the Tory party, had called for early elections expecting her party to gain further strength in Parliament. The results were different from these expectations, as Labour party gained 30 more seats in the new Parliament, while Tories, however, lost 13 seats [4].

8 DISCUSSION

We will now discuss results of our data analysis.

8.1 Analysis of US Election Tweets

We will first discuss results of subjectivity and polarity analyses performed on tweets gathered during the 2016 US elections.

8.1.1 Subjectivity Analysis. For our first research question, we performed subjectivity analysis of Twitter messages during the 2016 US elections. Here, we used Twitter location data to analyze which candidate was mentioned in more subjective tweets. Results of subjectivity analysis are shown in Figure 2. We can observe that tweets mentioning Donald Trump were more subjective than those mentioning Hillary Clinton. This phenomenon was observed across the board in all 10 states under study. We discovered a similar trend for all states of the US as shown in Figure 3.

Much research has been conducted regarding user behavior on Twitter. It is observed that a small number of users drive discourse on Twitter, resulting in lack of debate and divergence of opinions [6, 29]. In our subjectivity analysis, we have observed that tweets mentioning Donald Trump were more emotional and opinionated than factual. This high subjectivity was apparent during the election campaign, where a bitter and negative campaign was fought by both candidates with Donald Trump creating much controversy through his tweets [13].

A high subjectivity score does not indicate a higher propensity for a voter to vote. What it does indicate, however, is that Twitter users mentioning Donald Trump in their tweets were more opinionated.

8.1.2 Polarity Analysis. For our second research question, we performed polarity analysis for both presidential candidates in terms of Twitter location data. Polarity scores of both candidates for all 50 states are shown in Figures 4 and 5.

To perform a comparative analysis for both candidates, we calculated polarity for the 10 most populous states of the US. Results of polarity analysis for these states are shown in Figure 6 and in Table 4. Our polarity analysis was able to correctly gauge public sentiment in 8 out of the 10 states under study. However, for two states, Illinois and New York, where Hillary Clinton was victorious, our dataset shows Trump as having better sentiment among the Twitter users from those states. However, it should be noted that for both of these states the difference

between the polarity scores of the candidates is within the margin of error. Nonetheless, we state that our polarity analysis of Twitter location data remains inconclusive.

Twitter data analysis for election prediction has become an popular area of research. Many studies have claimed to effectively predict elections by determining public opinion utilizing Twitter data analysis [31, 38]. However, doubt has also been expressed by highlighting the limitations of utilizing Twitter data for election prediction [15, 27]. Our result using Twitter location data to gauge public sentiment towards candidates using Twitter is somewhat inconclusive. Out of the 10 states used for data analysis, sentiment of 8 was correctly predicted. Two states were predicted incorrectly, however, for both of these states the polarity scores of the two candidates were within the margin of error. We believe that further analysis of Twitter location data, where location is looked at in more detail, including city and county level, can generate interesting results for understanding public opinion towards political candidates.

8.2   Analyses of UK Elections Tweets

For the UK elections analyses, we downloaded two mutually exclusive datasets from Twitter. The purpose of this study was to test for consistency in results of our sentiment analyses. We evaluated sentiment of tweets in both datasets containing the words GE2017, Jeremy Corbyn, Theresa May, Labour, and Tory.

We can observe from Figure 7 that overall sentiment of the two datasets is dissimilar. This is verified by performing *t-test* on the sentiment values of the two datasets. However, for tweets containing the term **GE2017**, we observe similar sentiment in both datasets.

We observe similar results when evaluating sentiment towards candidates and parties in the two datasets. Figures 9 and 10 show sentiment of tweets towards candidates in the terms and location datasets, and we witness that, in both cases, Jeremy Corbyn enjoys better sentiment than Theresa May. This difference in sentiment is statistically significant and is verified through *t-test.*

Similarly, Figures 11 and 12 display sentiment of tweets mentioning Labour and Tory party and again in both cases, we observe Labour party having a statistically better sentiment than Tory in both results.

Although the trend of using Twitter sentiment analyses to gauge public opinion towards candidates and political parties during elections has grown of late, there exist many concerns regarding the use of this approach [15, 27]. One criticism in this regard is that results of Twitter election prediction studies cannot be replicated. By getting similar sentiment results from two different Twitter datasets, collected during the same time period for the same event, we believe that our results demonstrate the efficacy of using Twitter sentiment analyses to predict public opinion. By observing similar tendencies in the two datasets, we are able to generalize the findings of our election data analyses.

As discussed earlier, the Labour party performed better than expectations in the UK elections of 2017, and we observe this in our data analyses as well [3, 4].

9   CONCLUSION AND FUTURE WORK

In this research study, we perform sentiment analyses of Twitter location data gathered during the US and UK elections of 2016 and 2017, respectively. For US elections, we tested our research questions by extracting location data provided by users from tweet meta-data and use it to plot state-wise subjectivity and polarity on a map of the US. For UK elections, we gathered tweets using two different filters, keywords, and geo-location.

Overall, we observe sentiment based on location did reflect on ground public opinion. In UK elections, the Labour party performed better than expected and also had a more positive sentiment on Twitter according to our data analyses [3, 4]. In terms of subjectivity and polarity analyses of US election data, we observe tweets mentioning Donald Trump as having higher subjectivity than ones discussing Hillary Clinton. This was true for all 10 states under study. Polarity analyses gave us mixed results. Although Clinton had higher polarity than Trump in California, a state that she won by a wide margin, for the two other states that she won in the

elections (New York and Illinois), she had lower polarity score in our data analysis. However, scores in both cases were within margin of error. We believe that this ability to map user sentiment provides tremendous benefit in accurately predicting public opinion. Rather than evaluating all users as a whole, we are able to distinguish their opinions based on geographical location.

In the future, we would like to expand this study by performing more detailed and fine-grained location analyses. By dividing US states into rural and urban areas, we can further refine our location sentiment analyses as rural; urban divide plays a very important role in elections. With location-based sentiment analyses, opinions of citizens towards political candidates can be measured with greater accuracy.

We would also like to expand this study to other countries of the world. As discussed in Section 5.1.1, we created a dictionary to map sentiment utilizing location information provided by Twitter users. Through expanding this dictionary to include states and regions from other countries of the world, we would be able to create a similar sentiment map.

REFERENCES

[1] Lorien C. Abroms and R. Craig Lefebvre. 2009. Obama's wired campaign: Lessons for public health communication. *J. Health Commun.* 14, 5 (2009), 415–423.

[2] Sitaram Asur and Bernardo A. Huberman. 2010. Predicting the future with social media. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT'10)*, Vol. 1. IEEE, 492–499.

[3] BBC. 2017. UK election. Retrieved from https://www.bbc.com/news/uk-wales-politics-40195154.

[4] BBC. 2017. UK election results. Retrieved from https://www.bbc.com/news/election/2017/results.

[5] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* O'Reilly Media, Inc.

[6] Javier Borondo, A. J. Morales, Juan Carlos Losada, and Rosa M. Benito. 2012. Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish presidential election as a case study. *Chaos Interdisc. J. Nonlin. Sci.* 22, 2 (2012), 023138.

[7] Lori Brainard and Mariglynn Edlins. 2015. Top 10 US municipal police departments and their social media usage. *Amer. Rev. Pub. Admin.* 45, 6 (2015), 728–745.

[8] Guido Caldarelli, Alessandro Chessa, Fabio Pammolli, Gabriele Pompa, Michelangelo Puliga, Massimo Riccaboni, and Gianni Riotta. 2014. A multi-level geographical study of Italian political elections from Twitter data. *PloS One* 9, 5 (2014), e95809.

[9] Nadya A. Calderon, Brian Fisher, Jeff Hemsley, Billy Ceskavich, Greg Jansen, Richard Marciano, and Victoria L. Lemieux. 2015. Mixed-initiative social media analytics at the World Bank: Observations of citizen sentiment in Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. In *Proceedings of the IEEE International Conference on Big Data (Big Data'15)*. IEEE, 1678–1687.

[10] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: A content-based approach to geo-locating Twitter users. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. ACM, 759–768.

[11] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Trans. Depend. Sec. Comput.* 9, 6 (2012), 811–824.

[12] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *J. Commun.* 64, 2 (2014), 317–332.

[13] Brett Edkin. 2016. Study: Trump Benefited From 'Overwhelmingly Negative' Tone Of Election News Coverage. https://www.forbes.com/sites/brettedkins/2016/12/13/trump-benefited-from-overwhelmingly-negative-tone-of-election-news-coverage-study-finds/#43ab79f93202.

[14] Emilio Ferrara and Zeyao Yang. 2015. Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Comput. Sci.* 1 (2015), e26.

[15] Daniel Gayo Avello, Panagiotis T. Metaxas, and Eni Mustafaraj. 2011. Limits of electoral predictions using Twitter. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence.

[16] Matthew Glassman, Jacob R. Straus, and Colleen J. Shogan. 2010. Social networking and constituent communication: Member use of Twitter during a two-week period in the 111th Congress. 66.

[17] Jennifer Golbeck, Justin M. Grimes, and Anthony Rogers. 2010. Twitter use by the US Congress. *J. Amer. Soc. Inf. Sci. Technol.* 61, 8 (2010), 1612–1621.

[18] Stefanie Haustein, Timothy D. Bowman, Kim Holmberg, Andrew Tsou, Cassidy R. Sugimoto, and Vincent Larivière. 2016. Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *J. Assoc. Inf. Sci. Technol.* 67, 1 (2016), 232–238.

[19] Sara Hofmann, Michael Räckers, Daniel Beverungen, and Jörg Becker. 2013. Old blunders in new media? How local governments communicate with citizens in online social networks. In *Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS'13)*. IEEE, 2023–2032.

[20] Vishal Kharde, Prof Sonawane et al. 2016. Sentiment analysis of Twitter data: A survey of techniques. *Arxiv Preprint Arxiv:1601.06971* (2016).

[21] Soonhee Kim and Jooho Lee. 2012. E-participation, transparency, and trust in local government. *Publ. Admin. Rev.* 72, 6 (2012), 819–828.

[22] Rob Kitchin. 2014. The real-time city? Big data and smart urbanism. *GeoJournal* 79, 1 (2014), 1–14.

[23] Nan Li and Desheng Dash Wu. 2010. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Dec. Supp. Syst.* 48, 2 (2010), 354–368.

[24] Ana Carolina E. S. Lima, Leandro Nunes de Castro, and Juan M. Corchado. 2015. A polarity analysis framework for Twitter messages. *Appl. Math. Comput.* 270 (2015), 756–767.

[25] David Lorenzi, Jaideep Vaidya, Basit Shafiq, Soon Chun, Neelima Vegesna, Zamil Alzamil, Nabil Adam, Seth Wainer, and Vijayalakshmi Atluri. 2014. Utilizing social media to improve local government responsiveness. In *Proceedings of the 15th International Conference on Digital Government Research*. ACM, 236–244.

[26] Laura McKenna and Antoinette Pole. 2008. What do bloggers do: An average day on an average political blog. *Pub. Choice* 134, 1 (2008), 97–108.

[27] Panagiotis T. Metaxas, Eni Mustafaraj, and Dani Gayo-Avello. 2011. How (not) to predict elections. In *Proceedings of the IEEE 3rd International Conference on Privacy, Security, Risk and Trust (PASSAT'11) and the IEEE 3rd International Conference on Social Computing (SocialCom'11)*. IEEE, 165–171.

[28] Anshul Mittal and Arpit Goel. 2012. Stock prediction using Twitter sentiment analysis. Retrieved from http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf).

[29] A. J. Morales, J. C. Losada, and R. M. Benito. 2012. Users structure and behavior on an online social network during a political protest. *Physica A: Stat. Mech. Applic.* 391, 21 (2012), 5244–5253.

[30] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley. 2013. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In *Proceedings of the International Conference on Web and Social Media (ICWSM'13)*.

[31] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International Conference on Web and Social Media (ICWSM'10)* 11, 122-129 (2010), 1–2.

[32] John H. Parmelee. 2013. Political journalists and Twitter: Influences on norms and practices. *J. Media Pract.* 14, 4 (2013), 291–305.

[33] Nitesh Sharma, Rachit Pabreja, Ussama Yaqub, Vijayalakshmi Atluri, Soon Chun, and Jaideep Vaidya. 2018. Web-based application for sentiment analysis of live tweets. In *Proceedings of the 19th International Conference on Digital Government Research: Governance in the Data Age*. ACM, 120.

[34] Lei Shi, Neeraj Agarwal, Ankur Agrawal, Rahul Garg, and Jacob Spoelstra. 2012. Predicting US primary elections with Twitter. Retrieved from http://snap. stanford. edu/social2012/papers/shi.pdf.

[35] Internet Live Stats. 2017. SentiStrength. Retrieved from http://sentistrength.wlv.ac.uk/.

[36] Internet Live Stats. 2019. Twitter Statistics. Retrieved from http://www.internetlivestats.com.

[37] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *J. Assoc. Inf. Sci. Technol.* 61, 12 (2010), 2544–2558.

[38] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International Conference on Web and Social Media (ICWSM'10)* 10, 1 (2010), 178–185.

[39] Alessandro Vespignani. 2009. Predicting the behavior of techno-social systems. *Science* 325, 5939 (2009), 425–428.

[40] Katrin Weller. 2016. Trying to understand social media users and usage: The forgotten features of social media platforms. *Online Inf. Rev.* 40, 2 (2016), 256–264.

[41] Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computat. Ling.* 30, 3 (2004), 277–308.

[42] Ussama Yaqub, Soon Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2017. Sentiment-based analysis of tweets during the US presidential elections. In *Proceedings of the 18th International Conference on Digital Government Research*. ACM, 1–10.

[43] Ussama Yaqub, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2017. Analysis of political discourse on Twitter in the context of the 2016 US presidential elections. *Gov. Inf. Quart.* 34, 4 (2017), 613–626.

[44] Ussama Yaqub, Nitesh Sharma, Rachit Pabreja, Soon Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2018. Analysis and visualization of subjectivity and polarity of Twitter location data. In *Proceedings of the 19th International Conference on Digital Government Research: Governance in the Data Age*. ACM, 67.

[45] Xiaohua Zeng and Liyuan Wei. 2013. Social ties and user content generation: Evidence from Flickr. *Inf. Syst. Res.* 24, 1 (2013), 71–87.