# 

**Abstract-With resounding political issues such as Brexit and the increase in social media usage during the past decade, there’s a growing demand for unique approaches to predicting outcomes for political elections worldwide. As demonstrated through various related works, data from social media platforms such as Twitter, can be extrapolated using data mining techniques in the field of social network mining. Our team contributes to modern research in social network mining through the combined use of various social network mining techniques. Specifically, we analyzed the sentiment of various tweets collected in the United Kingdom’s major metropolitan areas during the 2019 federal election period. Using a time fading model, weighted graph and the AFINN-111 word list, our conclusion accurately predicted the political sentiment of the polls with respect to the average age of a Twitter user. Further refinements to our methodology such as a wider radius on our collection zones, would yield more universally accurate results. Our methodology can be expanded to predict the results of future elections as well, including the upcoming American presidential election in 2020.**

**Keywords-Election, Sentiment, Time Fading, Weighted, Batch, Node**

# **INTRODUCTION**

From telecommunications, healthcare, and business, the demand for predictive analytics has risen dramatically in the 21st century [1]. International politics is one of the largest areas hit by our increased access to data and those who wish to launch political careers can no longer ignore the way in which such technology and political life interact with one another [2]. With an increasingly divisive global political landscape and the resurgence of economically critical issues such as Brexit, political parties across the globe have begun to focus on data driven campaigns [3]. The growth of social media and the heightened interconnectivity of the modern world have effectively changed the methods by which political pundits, campaign personnel, and ordinary citizens alike predict the results of federal elections. In the past, political campaign focuses were driven in large by macro trends through historical data. Parties would focus on swinging regions over individuals [3]. Alongside the rise of computer technology in the late 20th century came the evolution of Data mining which can be defined as the process of analyzing data in novel ways to find unique relationships [7]. Now, political campaigns and pollsters can attempt to better analyze individual data to achieve more accurate predictions.

Known as the non-trivial extraction of implicit, previously unknown, and potentially useful information from data, data mining and its techniques including those with respect to similarity matrices and clustering have been used in recent years to enhance the predictions of the world’s democratic processes [4]. However, our team decided to take a more modern approach to predicting election data using social network mining (or social media mining). The idea is to use social media platforms and their breadth of data on an individual to determine their support for various political parties. When people think of social media mining, they might recall the Cambridge Analytica incident in which the firm harvested personal information from millions of users on Facebook to assist the Republican party in the 2016 presidential campaign [5]. The best part of social network mining is that it is typically free. Traditionally, pollsters would have to come up with a poll and distribute it to the public, modern day analysis allows few limitations, restricted primarily by the range of acceptable topics permissible on such platforms. With the economic ramifications of decisions such as Brexit, our team decided to conduct our own experiment with social network mining for the 2019 federal election in the United Kingdom.

The aim of this report is to observe existing models for the extraction of data from social networks and apply snippets of our findings on Twitter to predict the support for each of the parties in the election. Specifically, our team chose a combination of sentiment analysis and stream-based mining techniques to mine information from popular social media platform Twitter. While many similar experiments have been conducted in the past with respect to mining Twitter information, stream-based data mining, and sentiment analysis, our contribution is to amalgamate these techniques and present an effective alternative to mainstream polling. Before we decompose our methodology, we’ll first examine the works that inspired our techniques and research behind them.

# **II. RELATED WORK**

Popular social media platform Twitter where users post short length messages describing their current thoughts, currently has over 300 million active users [8]. Each individual message is referred to as a tweet. The study [27] examined tweets using the Twitter API (3.1), to predict the results of the 2008 United States presidential election. They assessed each tweet’s emotional sentiment as being either positive or negative. To do this, the team filtered the stream of tweets they collected into those referring to relevant topics such as the economy, jobs, McCain, and Obama. They analyzed the words in each tweet using the Opinionfinder database to assign a positive or negative score to each word. Then divided the total positive score by the total negative score to get the final sentiment of each tweet (1).

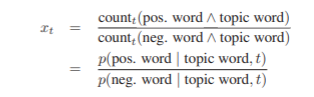


Figure 1 [9]

They concluded that the accuracy of their algorithm relied heavily on the topic of each batch of tweets. Those relating to Obama were more accurate than those relating to economy but even so the results weren’t perfect with statistical r values (a measure of inherent correlation) between 0.5 and 0.8 for the best of topics. It’s worth noting however, that this paper predicted results for an election 10 years ago when fewer individuals had access to social media.

Even before social media platforms rose in popularity, early signs of social media mining and sentiment analysis were already present. Bo Pang and Lillian Lee from Cornell University ran a similar experiment in 2002. The pair worked on the analysis of movie reviews, analyzing the domain of online reviews available from popular review website IMDB. Unlike the previous work, the sentiment analysis consisted of training data with words chosen by research assistants instead of a well-known database. This is likely due to the age of the experiment, as it was at a time when such analysis was quite novel in nature. Once again both positive and negative sentiment values were realized from words contained within the reviews. As few sentiment analysis algorithms existed in 2002, the duo ran classic Naive Bayes, maximum entropy, and support vector methods to enhance their accuracy. In the end, their analysis was poor compared to traditional categorization without sentiment analysis, although support vector methods came close. But the team didn’t have access to the platforms around today and nowadays there are countless algorithms to defend the use of sentiment analysis.

Take Finn Arup Nielsen’s popular AFINN lexicon [16]. Looking back to Twitter for a moment, Nielson found that he could update the sentiment value of existing ‘word lists’ including the Affective Norms for English Words. His updated list contains over 3000 commonly used words and can be retrieved quickly for use in today’s research and compared against various topics. As shown in figure 2 below, Nielson’s AFINN lexicon found that technology generally has a more negative sentiment in news than sports does.

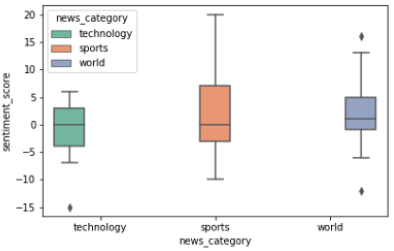


Figure 2: [16]

Another study conducted during the 2012 United States presidential election, analyzed tweets from Twitter concerning Barack Obama or Mitt Romney. Researchers [28] used the LIWC lexicon to compare which presidential candidate received more positive tweets. The study goes on to classify each tweet under the category of various economic issues such as taxation and healthcare to determine where the two ranked with respect to each issue. In order to do this, they gathered sub-categories (think of insurance as one for healthcare) for each major economic category from the tweets themselves. The candidates final score for each issue is based on the DPDT or the difference between positive and negative tweets for a particular issue.

Finally, our team considered research performed by Elder Santos during the 2014 Brazil general presidential elections [10]. While the related study on the Obama campaign [27] was released at a time when Twitter was only popular in the United States, today Twitter can be applied to sentiment analysis worldwide. Dilma Roussef won roughly 42% of the vote to become Brazil’s president after the election. Her runner-up received 34% of the vote [10]. With these results in mind, Santos began his analysis by using Twitter’s API to obtain tweets restricted to 14 cities in Brazil during the election. His setup included a graph with cities as the nodes and common trends as edges. Using a word list, he created, he counted the positive and negative supported trends along each edge in the graph. The results were astonishing, with an overall accuracy of 60% (see figure 3 below).

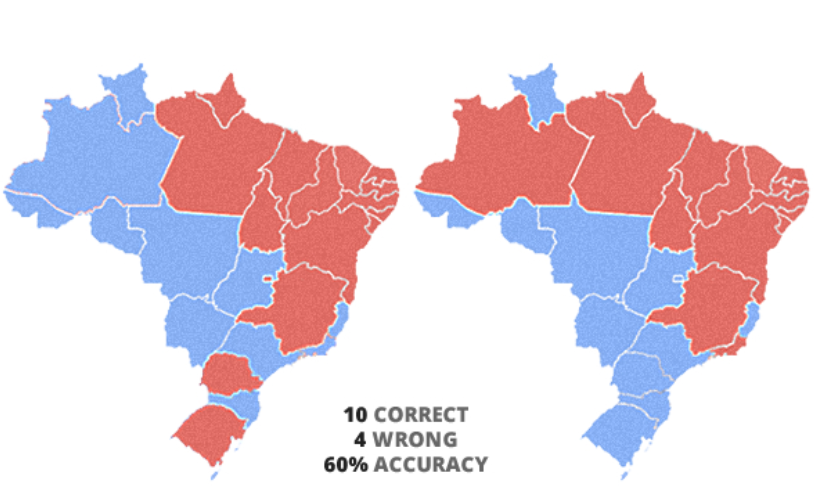


Figure 3: [10]

# **III. METHODOLOGY**

Our team purposes an alternative approach to traditional public opinion polling using a four-step process consisting of the data collection, sentiment scoring, network analysis, and support calculation phases.

**Data Collection**

Large batches of text-based information corresponding to individual users and their thoughts, feelings, and ideas can be gathered as tweets using the Twitter API (Application programming interface). Twitter’s API provides functionality for publishing, analyzing, and searching tweets [12]. Twitter’s API was an obvious choice for us due to its low learning curve and success in other works.

We consider the tweets for each day to represent one individual batch of data. The Twitter API gives us the list of tweets in a JSON format. Each tweet containing the tweet text body and the date. We then convert each JSON file to CSV (comma delimited) format for easy processing in further phases. Each batch has files containing tweets for all twelve cities (12 CSV files per batch). The CSV files contain five columns corresponding to the city for which the tweet was obtained, the respective political party for which the tweet body text is referring to. The leader of that party, the tweet body text, and finally the timestamp of the tweet. The name of each file represents the name of the data collection region (could be a city name, province, riding etc). For each tweet in every file, we save only the text body of the tweet which is a maximum of 115 characters as per the Twitter documentation [12].

**Sentiment Scoring**

For this step in the data analysis process, our team decided to take a different approach and make use of the AFINN-111 dictionary (section 2). Our reasoning for the selection of this word bank was due to the fact that the words contained within AFINN are specifically chosen based off research in social media. Only the most commonly used words on websites such as Twitter, are listed in this word list [16]. The AFINN lexicon assigns each word a sentiment value ranging from −5 (very negative) to +5 (very positive) using SentiStrength [16]. Modifications to this step, would allow for custom word lists as well.

Next, we create a hash map for the AFINN word list where the key in the key value pair (KVP) is the word and the value is the sentiment value. We recursively read every file contained within each batch and pass the tweet body into our ‘calculateSentiment’ algorithm. This algorithm takes in two parameters with the first being the tweet body text and the second being the AFINN HashMap. It then returns the sentiment value by taking the sum of all sentiment values for each AFINN word found in the tweet body and dividing it by the word count of each tweet. We divide specifically by the total word count and not just the word count for the AFINN words because tweets with fewer AFINN words should be weighed less accordingly. If a tweet does not contain as many AFINN words, our process shouldn’t let it override tweets that are more relevant to the lexicon we chose. This approach is strictly unique as related works such as Hu et al took the ratio of positive and negative sentiment values (figure 1), our team took the sum of each sentiment value assigned to the words in the tweet and divided it by the length of the tweets. We specifically chose this approach to limit the influence of shorter tweets on the overall sentiment. Generally, longer tweets are more comprehensible and convincing [16]. As such, they should be weighed more respectively. With the total sentiment value for all tweet body’s we create a new CSV file that is a partial copy of the original. The primary difference between the two however, is instead of containing the tweet body, it is replaced with the corresponding sentiment value. Finally, we use a function ‘isClean’ that can be set to handle custom parameters. Only tweets containing information relevant to these categories such as city, political party, etc. will be used. The purpose of this function is to verify the data prior to assigning it a sentiment score. It does this by comparing each of those three fields to ones found in an existing list we store within our application and tweets will be tagged with each of the chosen categories before they’re passed to the next stage in the process.

**The Network analysis**

Each of the tweets continue to be organized into batches based on time. Since the political sentiment may shift throughout election periods, our team incorporates a time fading approach for each batch in a separate trial run (in addition to an initial trial with no time fading). With the reading of each new batch, the results of older batches are multiplied by our alpha value α = 0.8. We chose an alpha value that represents the likely worst-case scenario in terms of sentiment shift during a typical election campaign. According to the findings of M. Ferell, this number is 17% or 0.17 which we round to 0.2 for easy math and additional confidence [24]. We then take 1.0 (100%) and subtract 0.2 (20%) to arrive at our alpha of 0.8.

At this point we construct a better method of organizing the data we collect. Using a network analysis, we create nodes labelled by each of the major political parties and regions of data collection for the respective election. These nodes are then connected by edges from one party node to one data collection region node. In addition, edges contain a number based on the total sentiment of each tweet relating to the nodes. The region nodes contain weights based on the political influence the region has in the democracy (i.e. number of seats). Each region weight W is determined using the following algorithm. We start by taking the total number of seats for the country denoted S. For each region X we divide the total population of the Country the region is in P by the population of our city C. Let σ = 0.01 in order to express each of our weights in terms of a percentage later. If we don’t reduce our result to a decimal value, we could end up with major cities (such as Beijing, Los Angeles, or London) having a weight on the graph above 1.0, which would eventually increase our weight (instead of decreasing it) when multiplied by percentage factors later. We multiply the resulting ratio by σ and then multiply that value by the number of seats S.

W=S\*((C/P)\*σ)

When creating the edges from a party to region we take the sentiment value given for the edge and multiply it by the weight of the region. The result is a positive or negative value l. This value is the sentiment towards a political party with respect to a specific data collection region. To determine how likely a party is to win the overall election we take the sum Σ(l1,l2,...,l12) or the total weights from all the edges for each party and add them together as Sentx (x representing one of the three political parties). For our time fading trial, this is done for each batch where the total value from all previous batches are multiplied by α (0.8, see above).

**Support Calculations**

Finally, we transform the sentiment scores for each party to percent support in order to obtain a more universal metric of analysis and easily compare our results to public opinion polls or related works. The sentiment scores are Sentp0 , Sentp1, … ,Sentpn. To calculate the total sentiment value, we do Senttotal = From p to n Σ|Sent|. Using those values we predict the support for each party Supp0 , ... , Suppn. Supparty = (Senttotal - Sentparty )/(2 \* Senttotal). This calculation method results in a combined support for all parties of 100%. We calculate support two ways, with and without time fading.

**Application to the United Kingdom Election**

We decided to gather all of our tweets in the areas surrounding the top three cities of all four countries in the UK (England, Scotland, Northern Ireland, and Wales). We opted to choose the top three in an attempt to strike a balance between the quality of our representative sample and the cost (both time and money) of acquiring more data.

Some cities are larger than others, so the area that tweets were drawn from varied depending on the population of the city. For each city a radius was specified, to determine the size of the area. The smallest of the twelve cities is Lisburn with a population of roughly 1/100th of the largest city, London. It was not desirable for the radius used for Lisburn to be 1/100th the size of London’s because that would be too small. To choose the radius for each city, the following calculation was used. First, cities were ordered from largest to smallest. Second, it was calculated ri = 50/i miles, where ri is the radius of ith largest city. Each day we gathered up to 100 tweets relating to each of the three major political parties in the UK, Conservative, Labour, and Liberal Democrat. Larger cities that are densely populated consistently pulled the full 100 tweet upper bound each day. The daily haul for smaller cities such as Lisburn varied depending on the number of relevant tweets sent during the 24-hour window of each day’s pull and ranged from 0-100. All of our tweets were gathered between November 18th to November 29th, 2019 or roughly two weeks out from the election date of December 12th, 2019.

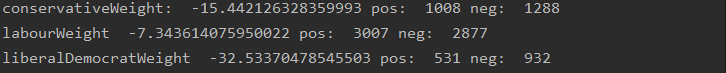
We run our data through the sentiment scoring procedure outlined in section 3.1 using the AFINN-111 word list as described. We explicitly filter tweets in the isClean() method with parameters city, party, and party leader, which make up columns in our new CSV file. Using the network analysis step above, we create nodes labelled by the three main political parties Conservative, Labour, and Liberal Democrat in addition to each of the twelve cities. We connect one party node to one city node and follow the procedure accordingly, with total number of seats in the United Kingdom clocking in at 650 [25] and the population of the country being 67,694,499 [26].

At this point, we calculate support using our first trial or simply the data produced without the use of time fading. The sentiment values for each party is SentLabour = -7.344, Sentcons = -15.442, and SentLiberal = -32.534. Senttotal = | -15.442 - 7.344 - 32.534 | = 55.32. SupLabour = (55.32 - 7.344)/(2\*55.32) = 43.362%. SupCons = (55.32 - 15.442)/(2\*55.32) = 36.043%. SupLiberal = (55.32 - 32.534)/(2\*55.32) = 20.595%. Without time fading our method puts Labour in first with 43.362% support, Conservatives in second with 36.043%, and Liberal Democrats with 20.595%.

Second, we calculate support for trial two that uses time fading. The sentiment values for each party is SentLabour = -0.491, Sentcons = -1.417, and SentLiberal = -5.844. Senttotal = | -1.417 - 0.491 - 5.844 | = 7.752. SupLabour = (7.752 - 0.491)/(2\*7.752) = 46.833%. SupCons = (7.752 - 1.417)/(2\*7.752) = 40.86%. SupLiberal = (7.752 - 5.844)/(2\*7.752) = 12.307%. With time fading our method puts Labour in first with 46.833% support, Conservatives in second with 40.86%, and Liberal Democrats with 12.307%.

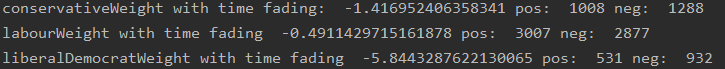
The results produced by our algorithm were initially quite shocking. After running each batch of data and coming up with a final result, our determination was as follows:

Figure 5: Without time fading



Without time fading, the conservative party has a negative sentiment value of -15.442, Labour has a sentiment value of -7.344, and the Liberal Democrats have a sentiment value of -32.534. All parties are viewed negatively overall. This shows Labour in first, second Conservatives, and third Liberal Democrats.

Figure 6: With time fading



With time fading, Labour is again in first with a sentiment score of -0.491, Conservatives in second with -1.417, Liberal Democrats in third with -5.844

# **IV. ANALYSIS**

So how did we do? Our team reviewed the findings of a publicly conducted metanalysis of local polls such as Ipsos and Delta poll [13-14]. As of December 5th, 2019, the latest public opinion polls were showing the Conservative party in the lead with an average support of 43%, followed by the Labour party with 32%, and finally the liberal democrats with an average support of 14%. Regardless of which trial we chose, our methodology puts the Labour party in the lead. So does this mean our process is entirely inaccurate? Not exactly. The difference between the public opinion polls and our results may be caused by the average age of Twitter users. According to Statista, the support of parties varies among age groups. The two lowest age groups measured were 18-34 and 35-44. We can assume that most Twitter users are in these two categories, because 62% of Twitter users are between the ages 18 and 49 [29].

Using the polling analysis from Statista, the support between 18-34 and 35-44-year-olds is Labour 33% and 34%, Conservatives 17% and 25%, and Liberal Democrats 13% and 13%. When we average the results after removing parties with under 10% support (which we didn’t consider for our algorithm) and undecided voters, Labour has 52.38% and 47.222% support, Conservatives have 20.635% and 34.722%, and Liberal Democrats have 20.635% and 18.056% support. Averaging the results gives 49.6%, 31.1%, and 19.5% support for Labour, Conservatives, and Liberal Democrats respectively for people between the ages of 18 and 44. This is close to the results our team produced with a difference of 6.2% for Labour, 4.9% for conservatives, and 7.2% for Liberal Democrats. All three of these results fall just outside or within the margin of error +/- 5%.

Applying time fading to the sentiment analysis suggests that the Conservative and Labour parties have gained support over time, while support for the Liberal Democrats has fallen.

Our data spans November 18th to 29th. Over that time period support for the Conservatives has increased by 3%, Labour has increased by 4%, and the Liberal Democrats have lost 1% of support. This is when comparing a Delta Poll survey [14] with an Ipsos poll [13].

As of December 17th, 2019, the results for the official election have been determined and the results are as follows. According to data published by the BBC, 44% of voters chose conservatives, 32% chose the Labour party, and 12% chose the Liberal Democrats [23]. Once again, this does not align with our results directly, however we can see that we are accurate once age is taken into account once again. Between the ages of 18-49 the Labour party finished the election with a whopping 48% support followed by the Conservative party at 29% support, and the Liberal Democrats at 13% support [22]. Our final accuracy was within 1.6% for Labour, 2.1% for the Conservatives, and lastly within 6.5% points for the Liberal Democrats.

While our team isn’t terribly disappointed with the results upon accounting for the age-related variable, there are several ways in which we can achieve far more accurate results upon future consideration of this work.

For one, we only accounted for the three main political parties in the United Kingdom. Initially we had planned on considering the top four, but we had to remove the Brexit party from our results due to interference in the naming of the party. Brexit is a critical issue and many of the tweets surrounding Brexit got pulled into our mining of Brexit party tweets. This distorted our data immensely. Conservative voters who are heavily in favour of Brexit, got lumped into the predictions for the Brexit party placing the Brexit party in a commanding lead each time. In the future we would take additional time to enhance our algorithm to account for this overlap.

In addition to the age oversight of Twitter users mentioned above, our team realized that many of the midland ridings outside of major metropolitan areas were neglected in the data collection phase. This was done to save time and ensure that a healthy supply of tweets were being mined each batch. However, it meant a heavy reduction in support for conservative voters, as according to the riding by riding breakdown of the public opinion polls, conservative garner a large percentage of their support outside of these major cities [23]. We can narrow down the results a lot more accurately by changing the gathering points for the data into smaller areas based on the actual neighborhoods of the ridings that will be voting for a seat. Rather than only using the 12 nodes (cities) we would have something closer to the actual number of seats available (650 nodes). As an example, with our present data, London is representing almost half the total ridings available. This does not accurately reflect its democratic representation as greater London accounts for only 73/650 seats [26]. Areas outside of London commanding their own riding, will have a much different outlook than in the city itself, yet we weigh them equally.

Furthermore, our team can use a more accurate lexicon or word list that’s specific to a political theme. While AFINN-111 works well for Twitter as a whole, it lacks many actual political terms with positive or negative sentiment such as ‘elected’ or ‘ouster’ respectively. This would likely enhance the accuracy of our sentiment values from a more political perspective.

Finally, we should ultimately consider gathering social media data from sources outside of just Twitter, to fairly represent the total population throughout all age groups. For instance, it has been shown that for adults over the age of 65, Facebook is the most popular social media website [9].

Finally, the techniques used above can be applied to many upcoming elections as well. Internet access continues to increase by 25 percent year over year in developing countries [20]. As more countries sign on to major social media platforms such as Twitter, elections results from civil war-torn countries with limited internal polling, could become easier to predict from across the globe. Even developed countries such as the United States can continue to benefit from this type of research as demonstrated by our first related work [27].

# **V. CONCLUSION**

With divisive political issues such as Brexit, the need for alternatives to mainstream public opinion polling has risen dramatically in recent years. As demonstrated by research concerning both the American and Brazilian federal elections, work around the world is being conducted to optimize the efficiency of mining data from social media platforms such as Twitter. Relevant tweets can be mined and assigned a sentiment value to determine the writer’s political affiliations. Universally accepted word lists such as the AFINN-111 can reduce the time and cost of research surrounding sentiment analysis but may ultimately limit results. Data streams of mined tweets can assist with filtering later in the mining process. For instance, using the timestamps of each batch, elections data can be gathered over a time period and weighted accordingly using the sliding window technique and the support of political parties can be predicted on a city by city basis using a weighted graph. In this paper, our team drew upon each of these techniques to produce an algorithm that can opt as an alternative to mainstream polling, when predicting elections results. Our efforts in predicting the results of the United Kingdom’s 2019 election simultaneously provides a glimpse into the possibilities of social network mining while demonstrating the need for further enhancements. Despite our results being far from what the overall polls are predicting, they are around the margin of error when accounting for the average age of Twitter users. It would behoove us to consider alternative social media platforms in an attempt to garner a more representative sample of data with respect to age. In a revised version of our experiment, our team would analyze data from more than just the 12 largest cities in the United Kingdom and focus on obtaining a more geographically representative sample of tweets. In addition, we would create our own proprietary word list and pair it with the existing AFINN-111 to obtain a more accurate sentiment value for each tweet. The methodology used above can be expanded to upcoming elections as well, including the 2020 American presidential election. Overall, we are happy with the results we’ve obtained, and our results suggest that we are not far off from competing directly with traditional polling. At the time of this writing, president Donald Trump tweeted an astonishing 113 tweets in a single day mostly in relation to ongoing impeachment proceedings with an increase in tweet volume on topics relating to the president as well [21]. As more and more election news and propaganda flood platforms such as Twitter, it increases the pool of data and overall accuracy of social network mining. One of the issues our team ran into, was finding quality batches of data in smaller towns. This is ultimately why we chose to predict within only the twelve major cities in the United Kingdom.

**REFERENCES**

|  |  |
| --- | --- |
| [1] | Chain of Demand, "Why Predictive Analytics Is Being Used By Managers," 3 April 2019. [Online]. Available: https://www.chainofdemand.co/the-power-of-predictive-analytics/. |
| [2] | J. Street, Politics and Technology, Macmillan, 1992. |
| [3] | Privacy International, "Case Study: Profiling and Elections - How Political Campaigns Know Our Deepest Secrets," 2019. [Online]. Available: Case Study: Profiling and Elections - How Political Campaigns Know Our Deepest Secrets. |
| [4] | W. B. T. M. L. P. a. H. B. A. Jakulin, "Data Mining in Politics," *Analyzing the U.S. Senate in 2003: Similarities, Clusters, and Blocs,* vol. 3, no. 17, pp. 291-310, 2009. |
| [5] | "Facebook and Data Privacy in the Age of Cambridge Analytica," University of Washington, 30 April 2018. [Online]. Available: https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/. [Accessed 14 Nov 2019]. |
| [6] | S. M. a. N. B. Dina Fawzy1, "The Evolution of Data Mining Techniques to Big Data Analytics: An Extensive Study with Application to Renewable Energy Data Analytics," *Asian Journal of Applied Sciences,* vol. 4, no. 3, 2016. |
| [7] | a. T. R. David W. Nickerson, "Political Campaigns and Big Data," University of Notre Dame, 2013. |
| [8] | J. Clement, "Statista," 3 August 2019. [Online]. Available: https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/. [Accessed 14 November 2019]. |
| [9] | R. B. B. R. R. a. N. A. S. Brendan O’Connor, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," in *Fourth International AAAI Conference on Weblogs and Social Media*, 2010. |
| [10] | E. Santos, "Data Mining for Predictive Social Network Analysis," Developers, 2018. [Online]. Available: https://www.toptal.com/data-science/social-network-data-mining-for-predictive-analysis. [Accessed 14 November 2019]. |
| [11] | D. Clark, "General election voting intention in the United Kingdom 2019 by age group," Statista, 8 November 2019. [Online]. Available: https://www.statista.com/statistics/1067740/uk-general-election-poll-by-age/. [Accessed 14 November 2019]. |
| [12] | Twitter, "Twitter Developer Labs," 2019. |
| [13] | Ipsos, "Ipsos MORI Political Monitor - November 2019," 2019. |
| [14] | DeltaPoll, "Deltapoll survey Results," DeltaPoll, 2019. |
| [15] | The London School of Economics and Political Science, "Social Media Platforms and Demographics," 2017. |
| [16] | F. A. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," DTU Informatics, Technical University of Denmark, Lyngby Denmark, 2011. |
| [17] | a. R. Dipanjan Sarkar, "Emotion and Sentiment Analysis: A Practitioner’s Guide to NLP," KD nuggets, August 2018. [Online]. Available: https://www.kdnuggets.com/2018/08/emotion-sentiment-analysis-practitioners-guide-nlp-5.html. [Accessed 14 November 2019]. |
| [18] | D. Levitt, "Election opinion polls tracker: gap between Labour and Tories narrows with result in balance," The Guardian, 12 December 2019. [Online]. Available: https://www.theguardian.com/politics/ng-interactive/2019/dec/05/election-polls-uk-2019-latest-opinion-poll-tracker. [Accessed 17 December 2019]. |
| [19] | L. A. Bowers, "Facebook, YouTube most popular social media among older adults," McKnights Senior Living, 23 July 2018. [Online]. Available: https://www.mcknightsseniorliving.com/home/news/facebook-youtube-most-popular-social-media-among-older-adults/. [Accessed 14 November 2019]. |
| [20] | J. M. J. B. M. C. a. A.-R. S. Olivia Nottebohm, "Online and upcoming: The Internet’s impact on aspiring countries," McKinsey&Company, 2012. |
| [21] | Associated Press, "Trump appears to hit new Twitter record with impeachment tweets," The Guardian, 13 December 2019. [Online]. Available: https://www.theguardian.com/us-news/2019/dec/13/trump-twitter-tweets-record-impeachment. [Accessed 17 December 2019]. |
| [22] | a. C. C. Adam McDonnell, "How Britain voted in the 2019 general election," YouGov, 17 December 2019. [Online]. Available: https://yougov.co.uk/topics/politics/articles-reports/2019/12/17/how-britain-voted-2019-general-election. [Accessed 17 December 2019]. |
| [23] | BBC, "UK results: Conservatives win majority," BBC, 17 December 2019. [Online]. Available: https://www.bbc.com/news/election/2019/results. [Accessed 17 December 2019]. |
| [24] | a. R. S.-B. David M. Farrell, Do Political Campaigns Matter?: Campaign Effects in Elections and Referendums, London: Routledge, 2002. |
| [25] | Worldmeters, "U.K Population," 2019. |
| [26] | M. Baxter, "General Election Prediction," Electoral Calculus, 2019. [Online]. Available: https://www.electoralcalculus.co.uk/homepage.html. [Accessed 17 December 2019]. |
| [27] | a. H. L. Xia Hu, "Mining and Profiling in Social Media," *SemanticsScholar,* 2012. |
| [28] | L. S. B. a. X. H. zz Amir Karami, "Mining Public Opinion about Economic Issues: Twitter and the U.S. Presidential Election," University of South Carolina, 2018. |
| [29] | Omnicore, "Twitter by the Numbers: Stats, Demographics & Fun Facts," Omnicore Group, 5 September 2019. [Online]. Available: https://www.omnicoreagency.com/twitter-statistics/. [Accessed 17 December 2019]. |