Lexicon-based Bot-aware Public Emotion Mining and Sentiment Analysis of the Nigerian 2019 Presidential Election on Twitter

Temitayo Matthew Fagbola[1], Surendra Colin Thakur[2]

Department of Computer Science, Federal University, Oye-Ekiti, PMB 373, Ekiti State, Nigeria[1]
e-Skills coLab, Durban University of Technology, Durban 4000, South Africa[1, 2]

Abstract—Online social networks have been widely engaged as rich potential platforms to predict election outcomes' in several countries of the world. The vast amount of readily-available data on such platforms, coupled with the emerging power of natural language processing algorithms and tools, have made it possible to mine and generate foresight into the possible directions of elections' outcome. In this paper, lexicon-based public emotion mining and sentiment analysis were conducted to predict win in the 2019 presidential election in Nigeria. 224,500 tweets, associated with the two most prominent political parties in Nigeria, People's Democratic Party (PDP) and All Progressive Congress (APC), and the two most prominent presidential candidates that represented these parties in the 2019 elections, Atiku Abubakar and Muhammadu Buhari, were collected between 9th October 2018 and 17th December 2018 via the Twitter's streaming API. *tm* and *NRC* libraries, defined in the '*R*' integrated development environment, were used for data cleaning and preprocessing purposes. Botometer was introduced to detect the presence of automated bots in the preprocessed data while NRC Word Emotion Association Lexicon (EmoLex) was used to generate distributions of subjective public sentiments and emotions that surround the Nigerian 2019 presidential election. Emotions were grouped into eight categories (sadness, trust, anger, fear, joy, anticipation, disgust, surprise) while sentiments were grouped into two (negative and positive) based on Plutchik's emotion wheel. Results obtained indicate a higher positive and a lower negative sentiment for APC than was observed with PDP. Similarly, for the presidential aspirants, Atiku has a slightly higher positive and a slightly lower negative sentiment than was observed with Buhari. These results show that APC is the predicted winning party and Atiku as the most preferred winner of the 2019 presidential election. These predictions were corroborated by the actual election results as APC emerged as the winning party while Buhari and Atiku shared very close vote margin in the election. Hence, this research is an indication that twitter data can be appropriately used to predict election outcomes and other offline future events. Future research could investigate spatiotemporal dimensions of the prediction.

Keywords—*Nigeria; 2019 presidential_election; bots-awareness; EmoLex; lexicon_analysis; public_opinion; emotion_mining; sentiment_analysis; twitter; APC; PDP; win_prediction; muhammadu_buhari; atiku_abubaka*

I. INTRODUCTION

The online social networks, being a medium for communicating and sharing opinions, have provided us with a large and rich variety of facts, interests and opinions which can be accessed and mined, by leveraging on the strong predictive power of learning algorithms, to generate useful patterns about offline events [1] [2]. Examples of such platforms include Twitter, Facebook, Instagram, YouTube among others. David and Francesco [3] defined opinion mining as a sub-discipline of computational linguistics that major on extracting people's opinion on social platforms. In recent years social media platforms have expanded massively thereby encouraging users to participate, contribute, and give opinions regarding different kind of issues, debates, and election prediction. Social platforms such as Twitter and Facebook provide researchers with a large significant data to be processed and mined due to the massive user-generated content which has widen the application scope of public opinion mining tools, and are becoming highly available to majority of the public [4]. Several works have adopted and utilized the readily-available data on social media platforms such as Twitter as a relatively cheaper and less-time consuming means for mining public opinions [5]. Public opinion mining comes handy and cogent in studies such as election predictions [6], crime analysis [7], Stock market analysis [8], consumers' behaviour [9], the outbreak of disease [10], public health [11] and so on.

In elections, opinion polls have been the intermediary between the opinion of the voters and politicians. However, the desire to predict an election outcome via twitter is to create an alternative to current polls with the minimal cost but still maintaining the standard in terms of precision, accuracy and reliability. Examples of elections forecasted via the use of online social networks include the 2018 Brazilian presidential elections [12], 2016 United States presidential election [13-16], Indian 2015 election [17], 2014 Brazil presidential election [18], Pakistan Election 2013 [19] and United States presidential election 2012 [20] among others. However, most of these works did not take into cognizance, the presence of software bots in online social networks. Software bots are automated accounts usually deployed by a third party to

pollute the content, manipulate and influence public sentiment in online social networks and performing human-associated activities like tweeting, retweeting and liking among others [21-22]. Zi, Steven, Haining and Sushil [23] revealed that Twitter accounts originating from human, cyborg and bot, follow an estimated 5:4:1 ratio distribution, respectively. Hence, the need for bots detection and removal from Twitter data as crucial preprocessing steps cannot be overemphasized.

Currently, the Nigeria 2019 presidential election is topical and of great interest to the public as it attracts diverse opinions that could be mined and analyzed to infer public preference in comparison with the actual outcomes of the election. In this paper, a bot-aware public emotion mining on Twitter is conducted to predict winner of the Nigeria's 2019 presidential election. This prediction leverages on the NRC Word Emotion Association Lexicon (EmoLex) [24] and an enhanced Sentiment Polarity Lexicon Model (SPLM) [25] to detect and analyze the distribution of sentiments (positive and negative) and emotions (happy, sadness, trust, anger, anticipation, fear, joy, surprise, disgust), as defined in the Plutchik's wheel of emotions [26], which are associated with the public opinions expressed on Twitter. The rest of the paper is organized as follows: Section 2 presents some related works regarding the use of online social networks for predicting presidential elections' results; Section 3 presents the materials and methods used to conduct the study; Section 4 presents the results of the analyses and predictions of the winner of the presidential election while the conclusion is presented in Section 5.

II. RELATED WORKS

Sri, Arash and Christine [27] analyzed the public sentiments of over 200,000 tweets associated with the two major candidates in the United States' 2016 Presidential Election in terms of positive, negative, anger, trust, anticipation, disgust, joy, fear, sadness and surprise. Results reported indicate a significant difference in positive sentiment and emotions including fear, joy, disgust, surprise, trust while negative sentiment and anticipation, anger and sadness emotions were reported insignificant. Brian, Joseph and Taghi [28] employed spatial information, positive sentiment and tweet volume features extracted from around 3 million tweets collected from twenty-one states between 22nd September 2016 and 8th November 2016 to predict the winner of the U.S. 2016 Presidential Election between the two prominent candidates "Donald Trump" and "Hillary Clinton". Budiharto and Meiliana [29] conducted a sentiment analysis and counting of tweets gathered between March 2018 and July 2018 with regards to the Presidential Candidates of Indonesia (Jokowi and Prabowo) to predict Indonesian Presidential election result. Andranik *et al.* [30] conducted a sentiment analysis of the 2009 federal election of the national parliament in Germany with over 100,000 tweets related to either a political party or a politician. The study affirmed that microblogging is able to provide extensive public opinion about offline events including politics. Gayatri and Nilesh [31] employed Naïve-Bayes and *k*-means algorithm to conduct a sentiment analysis about the public political orientation regarding the Maharashtra state assembly election by leveraging on the tweet contents and the associated twitter user-based features. Pritee and Sachin [32] adopted a lexicon-based approach integrated with Naïve-Bayes algorithm to detect and predict emotions in tweets using Gujarat Election data. Emotions of interest were tentativeness, sadness, anxiety, anger, work, certainty and achievement, as well as positive and negative sentiments. However, majority of the Candidates had more tweets expressing joy than any other emotion. However, the candidate Amit had the highest positive sentiment score, followed by Ahmed and Smriti in that order. Elvyna and Yustinus [33] applied Naïve-Bayes algorithm to predict the 2016 U.S. Presidential election with 371,264 tweets gathered between December 16, 2015 and February 29, 2016 using Twitter Streaming API on Tweepy. The system achieved a 95.8% accuracy over a 10-fold cross validation and also predicted Ted Cruz and Bernie Sanders as Republican and Democratic nominee respectively. Kassraie, Modirshanechi and Aghajan [34] developed a Gaussian process regression model to predict the 2016 US Election from 370,000 tweets via an estimation of the candidates' vote shares.

Adam and Alan [35] combined sentiment analysis with volume-based metrics to model political sentiment of the 2011 Irish General Election using data obtained from Twitter. The developed integrated method was evaluated against the conventional election polls and the final election result. Results obtained reveal that the developed method is predictive and capable of revealing public opinions about elections via social media data. Oikonomou and Tjortjis [36] employed Twitter data to predict the winner of the 2016 USA presidential elections in Ohio, Florida and North Carolina. To achieve this, the study conducted a sentiment analysis on the two major candidates "Donald J. Trump" and "Hillary Clinton" to predict the winner. Kellyton *et al.* [12] identified the correlation between social media and the final outcome of the 2018 Brazilian presidential elections for the thirteen (13) candidates using 291 million social media users' interactions and over 41,000 posts obtained from Instagram, Twitter and Facebook between January 2018 and October 2018. Features of interest adopted include social media profiles, users' posts and number of followers and the posts from the presidential candidates. The study established a strong correlation between the followers of candidates and votes received. Similarly, there was a recorded higher rate of discussions via posts and followers on Instagram than were recorded on both Facebook and Twitter.

Omaima, Suhem and Bravim [37] employed location information from 650,000 tweets collected from Twitter within a period of 5 days to identify patterns and trends in a bid to predict the winner of the Indian general elections 2014. The study used sentiment analysis and Naïve-Bayes algorithm for prediction and classification purposes. Jayakumar *et al.* [38] adopted a Long Short Term Memory (LSTM) deep learning model to predict wins in Indian elections by leveraging on users' opinions in twitter. Barkha and Sangeet [39] developed a Hybrid Topic Based Sentiment Analysis (HTBSA) that leveraged on biterm topic model and geotagging to extract latent topics and capture word relations and co-occurences in more than 300,000 tweets collected between 1st February 2017 and 20th February 2017 to predict

win in Uttar Pradesh legislative elections. Manish, Anoop and Amit [40] predicted the winner of the three presidential elections held in Latin America during the months of February through April, 2013 by leveraging on the popularity of the candidate on Twitter. Vinay and Shishir [41] developed an intelligent model based on dynamic keywords and topic modeling to predict win in elections with respect to the sentiments of users and volume of tweets. Chaitanya [42] conducted a predictive sentiment analysis of the Karnataka State Elections with tweets fetched obtained from Twitter using Tweepy. Zhang [43] employed regression analysis to develop a model to forecast the results of Taiwanese elections using data obtained from candidates' Facebook posts and their Like Ratio and daily average page views on each candidate's Wikipedia page. Bruno *et al.* [44] developed a spatio-temporal sentiment framework aided with a Support Vector Machine to analyze the Brazilian presidential elections using data obtained from Twitter. Di-Fatta *et al.* [45] conducted a political sentiment analysis of the 56th United Kingdom Parliament general elections with over 28 million tweets collected between 1st March 2015 and 31st May, 2015.

III. Materials and Method

In this section, the profiles of the political parties and that of the candidates, the description of the data collected and the Lexicon-based bot-aware election prediction workflow used for the predictive analysis are presented.

A. Materials

The Political Parties: The two (2) political particles are the All Progressive Congress (APC) and the People's Democratic Party (PDP).

1) Profiles of the political parties

a) APC: APC became the largest political party in Nigeria by toppling the PDP after an emphatic win at the national level in the 2015 Presidential election which brought in President Muhammadu Buhari. Apart from the presidency, APC has 22 states in its coffers as well as the president of the senate and the speaker of the House of Representatives. The leader of the party is Senator Bola Ahmed Tinubu.

b) PDP: PDP is the second biggest political party in Nigeria after losing the 2015 presidential election to the All Progressive Congress (APC). But it is still the most formidable opposition party to the ruling party. PDP still has about 12 states in its control as well as the deputy senate president and a handful of senators and honorable members.

2) Profiles of the presidential candidates:
The profile of the two prominent Nigerian 2019 presidential candidates is presented in Table I.

3) Inferring public voting intentions and preference from tweets:
A modified SPLM developed by Davidov, Tsur and Rappoport [25] is an excellent tool to infer and estimate public sentiment in online social networks about objects and/or offline social events. In this paper, SPLM is used to evaluate actual public voting preference regarding the political parties and the presidential candidates in the Nigerian 2019 presidential elections as presented in equations (1 and 2).

$$(P_1) = \frac{Pos(P_1)+Neg(P_2)}{Pos(P_1)+Neg(P_1)+Pos(P_2)+Neg(P_2)} \tag{1}$$

$$(P_2) = \frac{Pos(P_2)+Neg(P_1)}{Pos(P_1)+Neg(P_1)+Pos(P_2)+Neg(P_2)} \tag{2}$$

where P_1 and P_2 are the political parties, $Pos(P_1)$ depicts the positive sentiment score for P_1, $Pos(P_2)$ depicts the positive sentiment score for P_2, $Neg(P_1)$ depicts the negative sentiment score for P_1 while $Neg(P_2)$ depicts the negative sentiment score for P_2.

B. Lexicon-based Bot-aware Election Prediction Workflow

The workflow is comprised of four steps including data acquisition, data cleaning, preprocessing, tweet analysis and output visualization as presented in Fig. 1.

1) Data acquisition: 224,500 tweets associated with some leading political hashtags like #PDP2019, #APC2019, #Buhari2019, #Atiku2019, #MBuhari2019, #AtikuObi and #Atikulated2019 were fetched from Twitter via the Twitter REST API between 9th October 2018 and 17th December 2018. An example of the code snippet used to generate tweets from twitter for #Buhari2019 is as follows:

$Buhari <- searchTwitter("\#Buhari2019", n = 20000)$ where n represents the number of Tweets to be returned. Sample tweets collected are presented in Fig. 2. The data was collected into an Excel spreadsheet and saved as a CSV (comma separated values) format as shown Fig. 3. Different files were used to store tweets of each unique party and presidential candidate.

TABLE. I.   Information Regarding the Presidential Candidates

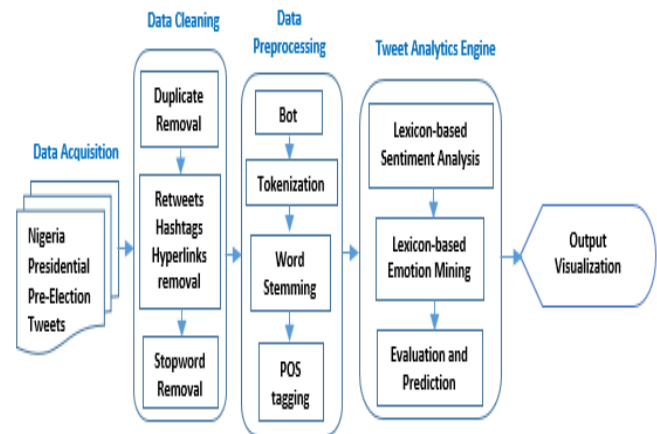| S/N	Full Name	Age (Years)	Position held
1.	Muhammadu Buhari	75	Incumbent President of the Federal Republic of Nigeria, he was also the Military Head of State from 1983 – 1985
2.	Abubakar Atiku	71	Former vice president of Nigeria

Fig. 1.   Our Lexicon-based Bot-aware Election Prediction Architecture.

RT @senatorsolace: If MUSA can score 2 goals in 2nd half, that means BUHARI can do better in 2nd term
@fkeyamo There is nothing distasteful in this, Keyamo is doing a great job, he is really giving opposition sleeples...
RT @BornoYoung: @MBuhari & @ProfOsinbajo will win the 2019 election wether you like it or not. #Buhari2
@MBuhari & @ProfOsinbajo will win the 2019 election wether you like it or not. #Buhari2019 C16
@APCNigeria Asiwaju, is a colossal personality and politician. Buhari2019…
@APCNigeria He's a Wise Choice. Man Of The People. Buhari2019…
@APCNigeria Cheerful News Please Do More. Buhari2019…

Fig. 2. Sample Tweets Collected Mentioning #Buhari2019.

Fig. 3. Cross-Section of the Tweets Collected in Microsoft Excel Spreadsheet.

2) Data cleaning: Data cleaning is an important step to reduce the noise in the dataset. Duplicate hashtags, punctuations, tweets, retweet entities, upper to lower case conversion, hyperlinks, numbers, stopwords, emojis, empty spaces, special and repeated characters in a text and HTML twitter handles were removed by introducing appropriate arguments into the "*gsub* ()" and the "*grep* ()" functions defined in the "*nrc*" documentation. Out of the 224,500 tweets fetched, 10,523 tweets/retweets were found to be duplicates and so removed. Sample '*R*' code snippet used to process tweets collected from twitter is as follows:

$$removeURL \; < -function(x) \; gsub \; ("http[^[:space:]] \\ *", "", x))$$

3) Data pre-processing: Data preprocessing steps include bot removal, tokenization, word stemming and part of speech tagging. The Botometer [46] was used to check for likely presence of bots in the remaining 213,977 tweets. However, 45,244 tweets, with bot scores above 0.6, were detected to have originated from bot accounts and hitherto removed. After bot removal, the remaining tweets were tokenized into a set of unigram words using the *get_tokens* () function defined in the Syuzhet library in '*R*'. Furthermore, each unigram is reduced to its shortest possible form using the *stemDocuments* () function defined in the "*tm*" library in *R*. For example, "voted" becomes "vote". Moreover, POS tagging was conducted to disambiguate the meaning of the unigrams and their lexical categories. Tables II and III present the common hashtags used by the two (2) presidential candidates and the two (2) dominating political parties on Twitter alongside their respective number of remaining after preprocessing.

TABLE. II. 2019 NIGERIAN PRESIDENTIAL ELECTION'S CANDIDATES, HASHTAGS AND TWEETS AFTER PREPROCESSING

Candidates	Search Keywords (Hashtags)	# tweets
Muhammadu Buhari	#Buhari2019, #MBuhari2019	24,897
Abubakar Atiku	#Atiku2019, #AtikuObi2019, #Atikulated2019	26,354

TABLE. III. 2019 NIGERIAN PRESIDENTIAL ELECTION'S PARTIES, HASHTAGS AND TWEETS AFTER PREPROCESSING

Parties	Search Keywords (Hashtags)	# tweets
All Progressive Congress	#APC2019, #BuhariOsinbajo2019	86,659
People's Democratic Party	#PDP2019, #AtikuObi2019	30,823

4) Tweet analytics engine: The analytics engine employs the Syuzhet library functions defined within the '*R*' documentation and the NRC Word Emotion Association Lexicon (EmoLex) dictionary to estimate the scores of the eight (8) Plutchik's emotion distributions (sadness, trust, anger, fear, joy, anticipation, disgust, surprise) and the two (2) sentiments (negative and positive), associated with the tweets. However, a modified SPLM was used to infer and estimate public voting intention and win prediction by leveraging on the public sentiment scores.

IV. RESULTS

In this section, the public emotion distribution across political parties and the presidential candidates are presented. Similarly, results of public voting intention and preference analysis from the tweets are also presented.

A. Results of Public Emotion Distribution by Political Parties

1) All progressive congress: As presented in Fig. 4, the public emotion scores of APC towards the Nigerian 2019 Presidential election are 6077, 9836, 3150, 8487, 5039, 4906, 3087, 14316, 11527 and 20234 for "Anger", "Anticipation", "Disgust", "Fear", "Joy", "Sadness", "Surprise", "Trust", "Negative sentiment" and "Positive sentiment", respectively. Pictorially as shown in Fig. 5, it can be seen that most of the tweets conveyed "Positive" sentiment which accounted for about 23% of the entire NRC emotions distribution. "Trust" is the second most conveyed emotion and accounts for 16% of the overall emotions. "Negative" sentiment is the third most significant expression with a 13% distribution among other emotions. "Anticipation" (11%), "Fear" (10%), "Anger" (7%), "Joy" and "Sadness" (6%) while "Disgust" and "Surprise" has the least distribution with about 4% of the entire NRC emotions.
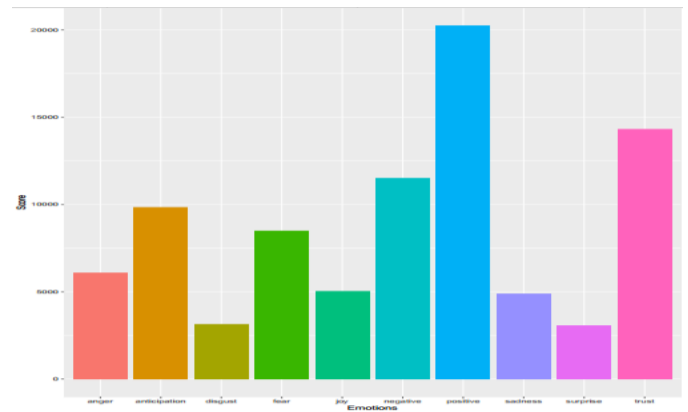

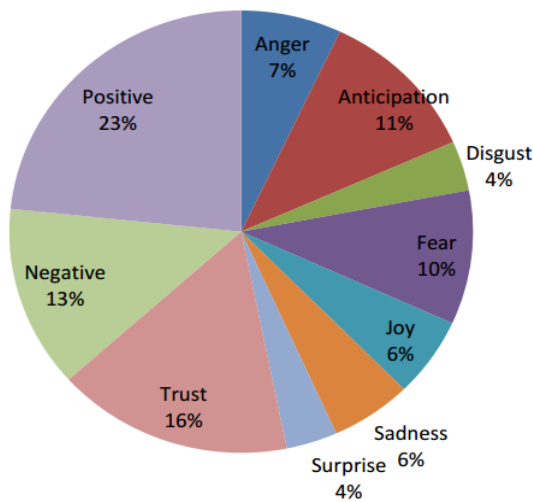Fig. 4. Chart Showing Public Emotion and Sentiment Scores for APC2019.

Fig. 5.   Chart Showing % Public Emotion and Sentiment Distribution for APC2019.

2) People's democratic party: As presented in Fig. 6, the public emotion scores of PDP towards the Nigerian 2019 Presidential election are 2236, 2944, 1521, 2587, 2291, 2166, 1334, 4737, 4753 and 6254, for "Anger", "Anticipation", "Disgust", "Fear", "Joy", "Sadness", "Surprise", "Trust", "Negative sentiment" and "Positive sentiment", respectively. The frequency of unique tweets depicted in percentage as presented in the pie chart of Fig. 7 shows that most of the tweets conveyed the "Positive sentiment" which accounted for about 20% of the overall distribution of the emotions, followed by "Negative sentiment" (16%), "Trust" (13%), "Anticipation" (10%), "Joy" (8%), "Fear" (8%), "Anger" (7%), "Sadness" (7%), "Disgust" (5%) and "Surprise" (4%) in decreasing order of public expression.

B. Results of Public Emotion Distribution by Presidential Candidates

1) #Buhari2019: The public emotion scores of "Buhari" towards the Nigerian 2019 Presidential election are 1672, 2405, 1104, 1958, 1842, 1717, 1086, 4099, 3682 and 5332 for "Anger", "Anticipation", "Disgust", "Fear", "Joy", "Sadness", "Surprise", "Trust", "Negative sentiment" and "Positive sentiment", respectively as presented in Fig. 8. A pie chart showing the percentage distribution of each of these emotions is presented in Fig. 9. It is observed that most of the tweets conveyed the "Positive sentiment" which accounted for about 21% of the overall distribution of the emotions, followed by "Trust" (17%), "Negative sentiment" (15%), "Anticipation" (10%), "Fear" (8%), "Anger" (7%), "Sadness" (7%), "Joy" (7%), "Disgust" (4%) and "Surprise" (4%) in decreasing order of public expression.

2) #Atiku2019: The public emotion scores of "Atiku" towards the Nigerian 2019 Presidential election are 1734, 2595, 1090, 1976, 2067, 1759, 1271, 4288, 3584 and 5990 for "Anger", "Anticipation", "Disgust", "Fear", "Joy", "Sadness", "Surprise", "Trust", "Negative sentiment" and "Positive sentiment", respectively as presented in Fig. 10. A pie chart

showing the percentage distribution of each of these emotions is presented in Fig. 11. It is observed that most of the tweets conveyed the "Positive sentiment" which accounted for about 23% of the overall distribution of the emotions, followed by "Trust" (16%), "Negative sentiment" (14%), "Anticipation" (10%), "Joy" (8%), "Fear" (7%), "Sadness" (7%), "Anger" (6%), "Surprise" (5%) and "Disgust" (4%) in decreasing order of public expression.
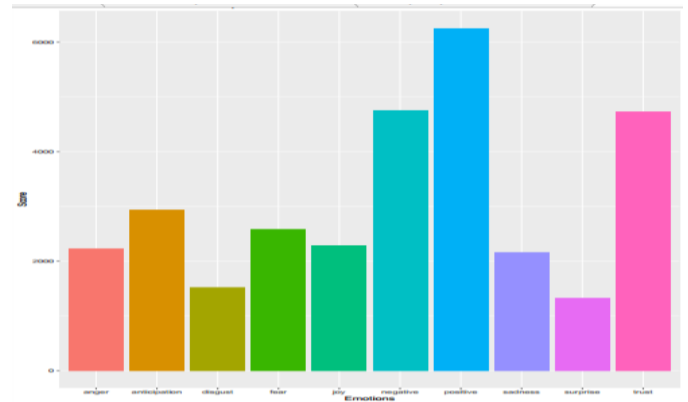

Fig. 6.   Chart Showing Public Emotion and Sentiment Scores for PDP2019.
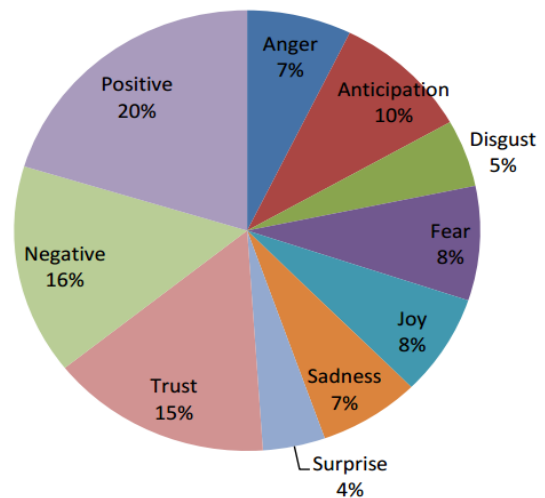

Fig. 7.   Chart Showing % Public Emotion Distribution for PDP2019.
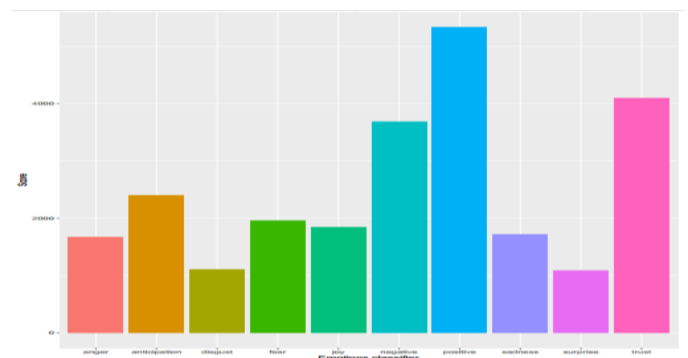

Fig. 8.   Chart Showing Public Emotion and Sentiment Scores for Buhari2019.
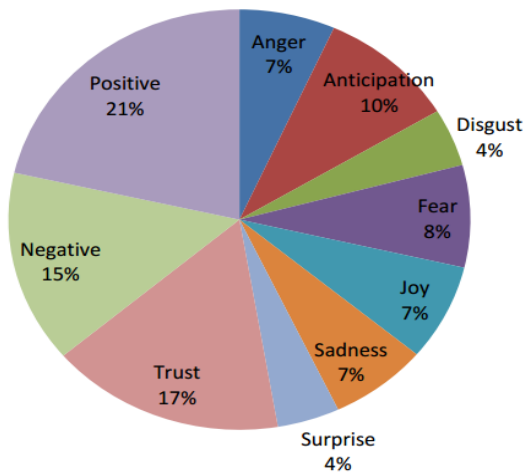
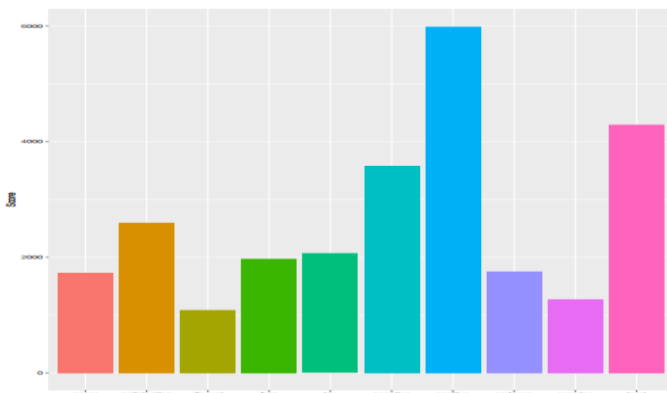Fig. 9.   Chart Showing % Public Emotion Distribution for Buhari2019.

Fig. 10. Chart Showing Public Emotion and Sentiment Scores for Atiku2019.
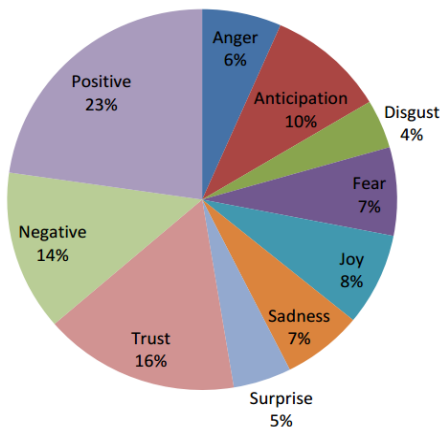

Fig. 11. Chart Showing % Public Emotion Distribution for Atiku2019.

3) Inferring public voting intention and preference from the tweets: The public voting preference for the political parties and the presidential aspirants from the analysis of the associated tweets is presented in this section.

4) Public opinion estimation of preferred political party: This was conducted to estimate the public preference towards a political party of choice in the 2019 Nigeria presidential election using a modified SPLM based on the party-based sentiment polarity scores obtained and presented in Table IV.

TABLE. IV.   PARTY-BASED SENTIMENT POLARITY SCORES

Description	Values
P_1	Party 1 (APC)
P_2	Party 2 (PDP)
$Pos(P_1)$	20234
$Pos(P_2)$	6254
$Neg(P_1)$	11527
$Neg(P_2)$	4753

Hence, owing to equations (1 and 2),

$$P(\text{APC}) = \frac{20234 + 4753}{20234 + 11527 + 6254 + 4753} = \frac{24987}{42768} = 0.58$$

$$P(\text{PDP}) = \frac{6254 + 11527}{20234 + 11527 + 6254 + 4753} = \frac{17781}{42768} = 0.42$$

Therefore, the public opinion as revealed by this analysis indicates that APC and PDP had a 58% and a 42% probability of win, respectively.

5) Public opinion estimation of preferred presidential candidate: The public opinion estimation of most preferred presidential candidate in the Nigerian 2019 presidential election was conducted for the two most dominant presidential candidates, (Atiku Abubakar and Muhammadu Buhari), using a modified SPLM that leverages on the sentiment polarity scores of the public expressed about the candidates as presented in Table V.

With respect to equations (1 and 2),

$$P(\text{Atiku}) = \frac{5990 + 3682}{5990 + 3584 + 5332 + 3682} = \frac{9672}{18588} = 0.52$$

$$P(\text{Buhari}) = \frac{5332 + 3584}{5990 + 3584 + 5332 + 3682} = \frac{8916}{18588} = 0.48$$

These results indicate that 52% of the public opinions expressed support for Atiku to emerge as the president of Nigeria while the remaining 48% goes to Buhari. Overall, the results of the party-based and the candidate-based sentiment analyses indicate that the public prefers APC to emerge as the ruling party but with some level of distastefulness for Buhari. The public had shown more interest in Atiku and not in PDP, his party. However, the outcome of the actual election held on the 23 February 2019, as reported by a widely-read newspaper, "The Guardian", with the headline "Muhammadu Buhari wins Nigerian election with 56% of the vote", declared Buhari as the presidential winner with 15.2 million votes over PDP's Atiku with 11.3 million votes. It was further stated that Buhari won in 19 out of the 26 states in Nigeria while Atiku won in the remaining 17 states. [1] These actual outcomes corroborate our predictions strongly as summarized in Table VI. These results have shown that online social networks is a veritable and effective tool for reliable

---

[1] www.theguardian.com/world/2019/feb/26/muhammadu-buhari-to-claim-victory-in-nigerias-presidential-election

assessment of public opinion regarding political issues and any other real-life events regardless of the domain of application. Furthermore, public opinion mining and sentiment analysis could help to evaluate public opinion and preference about an event and help make improved decisions.

TABLE. V.   CANDIDATE-BASED SENTIMENT POLARITY SCORES

Description	Values
C_1	Candidate 1 (Atiku)
C_2	Candidate 2 (Buhari)
$Pos(C_1)$	5990
$Pos(C_2)$	5332
$Neg(C_1)$	3584
$Neg(C_2)$	3682

TABLE. VI.   PREDICTED VERSUS ACTUAL WIN OUTCOMES IN THE NIGERIAN 2019 PRESIDENTIAL ELECTION

Parties / Candidates	Predicted Values (%)	Actual Outcomes (%)
APC / Buhari	58 / 42	56
PDP / Atiku	42 / 58	44

V. CONCLUSION

In this paper, a bot-aware, lexicon-based emotion mining and sentiment analysis of public opinions on Twitter were conducted to predict winner of the Nigeria's 2019 presidential election. 224,500 tweets, associated with the two most prominent political parties (PDP and APC) and the presidential candidates (Atiku Abubakar and Muhammadu Buhari), were collected between 9th October 2018 and 17th December 2018 via the Twitter's streaming API. *tm* and *NRC* libraries were used for data cleaning and preprocessing purposes. Botometer was introduced to detect the presence of automated bots in the preprocessed data while NRC Word Emotion Association Lexicon (EmoLex) was used to generate distributions of subjective public sentiments and emotions that surround the Nigerian 2019 presidential election based on the Plutchik's emotion wheel. Results obtained indicate a higher positive and a lower negative sentiment for APC than was observed with PDP, and as further corroborated by the actual election results. This strongly indicate that Twitter is a great resource for predicting offline social events. Future works could be directed towards incorporating both time and location metadata information from the tweets in the prediction.

REFERENCES

[1] R. Shah, and B. Patel, "Procedure of Opinion Mining and Sentiment Analysis," International Journal of Current Engineering and Technology, 4, pp. 4086-4090, 2014.

[2] C. Andrea, C. Luigi, I. Stefano, and P. Giuseppe, "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France," New Media & Society, vol. 16, pp. 340-358, 2013.

[3] O. David, and M. Francesco, "Research challenge on Opinion mining and sentiment analysis," Australian Journal of Public Administration, 64, pp. 3–13, 2012.

[4] N. Burns, Y. Bi, H. Wang, and T. Anderson, "Sentiment Analysis of Customer Reviews: Balanced versus Unbalanced Datasets," Knowledge-Based and Intelligent Information and Engineering Systems. Lecture Notes in Computer Science, vol. 6881(1), pp. 20-29, 2011.

[5] A. Karami, L. Bennett, and X. He, "Mining Public Opinion about Economic Issues: Twitter and the U.S. Presidential Election," International, IGI Global, vol. 9(1), pp. 18-28, 2018.

[6] N. Dwi, and C. Hauff, "Twitter-based Election Prediction in the Developing World," HT 15, Proceedings of the 26th ACM Conference on Hypertext & Social Media, pp. 149-158, 2015.

[7] W. Xiaofeng, G. Matthew, and B. Donald, "Automatic Crime Prediction Using Events Extracted from Twitter Posts," International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, 7227, pp. 231-238, 2012.

[8] B. Johan, M. Huina, and Z. Xiao-Jun, "Twitter mood predict the stock market," Journal of Computational science, 2(1), pp. 1-8, 2011.

[9] K. Evangelos, K. Areti, T. Efthimios, and T. Konstantinos, "Applying brand equity theory to understand consumer opinion in social media," Journal of Universal Computer Science, 22(5), pp. 709–734, 2016.

[10] R. Joshua, O. Miles, and K. Ewan, "Using prediction markets and Twitter to predict a swine flu pandemic," In Proceedings of the 1st International Workshop of Mining Social Media, pp. 9-17, 2009.

[11] M. Jonathan, "Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience," Journal of Elections, Public Opinion & Parties, vol. 24(1), pp. 45-72, 2014.

[12] K. Brito, N. Paula, M. Fernandes, and S. Meira, "Social Media and Presidential Campaigns – Preliminary Results of the 2018 Brazilian Presidential Election," 20th Annual International Conference on Digital Government Research, DOI:10.2245/3325112.3325252, 2019.

[13] V. Ignasi, T. Tina, and X. Wei, "Characterizing the 2016 U.S. Presidential Campaign using Twitter Data," International Journal of Advanced Computer Science and Applications, vol. 7(10), pp. 12–19, 2016.

[14] C. Delenn, Z. Anna, and Z. Jessica, "Analyzing Twitter Sentiment of the 2016 Presidential Candidates," SIAM Journal on Scientific Computing, vol. 38(4) pp.1-12, 2016.

[15] P. Burckhardt, "Tweet as a Tool for Election Forecast : UK 2015 General Election as an Example," IEEE Transactions on Knowledge and Data Engineering, pp. 1–47, 2016.

[16] J. Kazem, and M. Yumi, "The Predictive Power of Social Media: On the Predictability of U.S. Presidential Elections using Twitter," IEEE Transactions on Knowledge and Data Engineering, pp. 11-49, 2014.

[17] W. Gayatri, and A. Nilesh, "Indian election using Twitter," Advances in Intelligent Systems and Computing, vol. 468 (22), pp. 251–261, 2017.

[18] C. M. Carvalho, H. Nagano, and A. K. Barros, "A Comparative Study for Sentiment Analysis on Election Brazilian News," Proceedings of Symposium in Information and Human Language Technology, Uberlandia, MG, Brazil, October 2-5, 2017.

[19] R. M. Asif, Q. A. Mustafa, and B. H. Syed, "Prediction and Analysis of Pakistan Election 2013 based on Sentiment Analysis," ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, (December), pp. 700–703, 2014.

[20] S. Amith, A. Jadhav, P. Kapanipathi, C. Lu, H. Purohit, G. A. Smith, and W. Wang, "Twitris: A System for Collective Social Intelligence," In: Alhajj R., Rokne J. (eds) Encyclopedia of Social Network Analysis and Mining, Springer-Verlag New York, vol. 1, pp. 2240-2253, 2014.

[21] C. Shao, G. L. Ciampaglia, O. Varol, K. Yang, A. Flammini, and F. Menczer, "The Spread of Low-credibility by Social Bots," Nature Communications, 9, vol. 4787, pp. 1-9, 2018.

[22] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online Human-Bot Interactions: Detection, Estimation, and Characterization," Proceedings of the Eleventh International AAAI Conference on Web and Social Media. arXiv preprint arXiv:1703.03107, 2017.

[23] C. Zi, G. Steven, W. Haining, and J. Sushil, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?," IEEE Transactions on Dependable and Secure Computing, vol. 9(6), pp. 811-824, 2012.

[24] S. M. Mohammad, and D. P. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," Computational intelligence, vol. 29(3), pp. 1-25, 2013.

[25] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced Sentiment Learning using Twitter Hashtags and Smileys," COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: pp. 241-249, 2010.

[26] R. Plutchik, "The Psychology and Biology of Emotion," New York: Harper Collins, 1994.

[27] H. D. K. Sri, N. Arash, and W. Christine, "Identifying Trending Sentiments in the 2016 U.S. Presidential Election: A Case Study of Twitter Analytics," Issues in Information Systems, 18(2), pp. 80-86, 2017.

[28] H. Brian, D. P. Joseph, and M. K. Taghi, "Location-Based Twitter Sentiment Analysis for Predicting the U.S. 2016 Presidential Election," The Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31), pp. 265-270., 2018.

[29] W. Budiharto, and M. Meiliana, "Prediction and Analysis of Indonesia Presidential Election from Twitter using Sentiment Analysis," Journal of Big Data, vol. (5)51, pp. 1-10, 2018.

[30] T. Andranik, O. S. Timm, G. S. Philipp, and M. W. Isabell, "Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape," Social Science Computer Review, vol. 29(4), 402-418, 2011.

[31] P. W. Gayatri, and V. A. Nilesh, "Analysis of Indian Election using Twitter," International Journal of Computer Applications, vol. 121(22), pp. 37-41, 2015.

[32] S. Pritee, and D. Sachin, "Twitter Based Election Prediction and Analysis," International Research Journal of Engineering and Technology, vol. 4(10), pp. 539-544, 2017.

[33] T. Elvyna, and E. S. Yustinus, "And the Winner is ...: Bayesian Twitter-based Prediction on 2016 U.S. Presidential Election," IC3INA 2016 Conference, 3-5 Oct. 2016.

[34] P. Kassraie, A. Modirshanechi, and H. Aghajan, "Election Vote Share Prediction using a Sentiment-based Fusion of Twitter Data with Google Trends and Online Polls," In Proceedings of the 6th International Conference on Data Science, Technology and Applications (DATA 2017), pp. 363-370, 2017.

[35] B. Adam, and F. S. Alan, "On Using Twitter to Monitor Political Sentiment and Predict Election Results," Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology, IJCNLP 2011, pp. 2-10, 2011.

[36] L. Oikonomou, and C. Tjortjis, "A Method for Predicting the Winner of the USA Presidential Elections using Data extracted from Twitter. In the Proceedings of the 2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), pp. 1-8, vol. 2017.

[37] A. Omaima, P. Suhem, and C. Bravim, "Application of Location-Based Sentiment Analysis Using Twitter for Identifying Trends towards Indian General Elections 2014," IMCOM '15, January 08 - 10 2015, BALI, Indonesia.

[38] S. JayaKumar, P. Priyank, K. S. Rajat, and P. Akkshansh, "Election Result Prediction Using Deep Learning Techniques," International Journal of Engineering and Advanced Technology, vol. 8(4), pp. 74-77, 2019.

[39] B. Barkha, and S. Sangeet, "On Predicting Elections with Hybrid Topic based Sentiment Analysis of Tweets," 3rd International Conference on Computer Science and Computational Intelligence, Procedia Computer Science, pp. 346-353, 2018.

[40] G. Manish, K. Anoop, S. Amit, and B. Raytheon, "Leveraging Candidate Popularity on Twitter to Predict Election Outcome. ACM SNA-KDD '13 Chicago, Illinois, USA, pp. 1-8. 2013.

[41] K. J. Vinay, and K. Shishir, "Towards Prediction of Election Outcomes Using Social Media," I.J. Intelligent Systems and Applications, vol. 12, pp. 20-28, 2017.

[42] A. Chaitanya, "Predicting the results of the 15th Karnataka Legislative Assembly Elections using Twitter," International Journal of Scientific Development and Research, vol. 3(10), pp. 105-109, 2018.

[43] X. Zhang, "Social media popularity and election results: A study of the 2016 Taiwanese general election," PLoS ONE, vol. 13(11), 2018.

[44] J. G. P. Bruno, P. C. L. João, P. A. M. João, L. L. Fabio, T. Rafael, and B. P. Juliano, "Spatio-Temporal Trend Analysis of the Brazilian Elections based on Twitter Data," IEEE International Conference on Data Mining Workshops, pp.1355-1360, 2018.

[45] F. Di-Fatta, J. Reade, S. Jaworska, and A. Nanda, "Big social data and political sentiment: the tweet stream during the UK General Election 2015 campaign," In: The 8th IEEE International Conference on Social Computing and Networking (SocialCom 2015), Dec. 1921, 2015, Chengdu, China.

[46] N. Chavoshi, H. Hamooni and A. Mueen, "DeBot: Twitter Bot Detection via Warped Correlation," in ICDM, pp. 817-822, 2016.