AN EXPLORATION OF USING TWITTER DATA TO PREDICT THE RESULTS OF THE U.S. PRIMARY ELECTIONS

Jeffrey P. Kaleta, Georgia Southern University, jkaleta@georgiasouthern.edu
Hayden Wimmer, Georgia Southern University, hwimmer@georgiasouthern.edu

ABSTRACT
The use of social media user feeds is a common interest of researchers exploring public views and opinions. In this exploratory study, we look to investigate how Twitter feeds during a presidential primary election can be evaluated to determine the relationships between contesting candidates and garner any predictive insight into election contest outcomes. In this study we collect data from both the REST API and STREAMING API from Twitter, each having their own data collection merits, and perform an association analysis, sentiment analysis, and linear regression to determine what insights can be captured from the data. In this work we find revealing relationships between candidate users accounts on how they interact with each other. We also show how sentiment from verified user accounts on Twitter show significance in election contest outcomes.

INTRODUCTION
The role of Twitter in researching the population’s perceptions regarding common pop culture topics, what they do in their everyday life, and, as the focus of this study, how the public political perceptions serve as predictions toward election results, is a growing interest to data researchers. To investigate what can be done using data from Twitter, this study is an exploratory research project to investigate how can be used Twitter use for prediction of election results and how we can use that knowledge to move forward in the analysis of data through Tweets. More specifically, we look to answer the following questions:

1. Is social media a reliable predictor of public preferences in primary election outcomes? Furthermore, are there any significant indicators in tweet data that are more reliable in the prediction of U.S. voter based contests?
2. Do users of Twitter, as a social media platform, reflect the political landscape and public perceptions that affect political contest results? And can the political landscape be identified through users’ posts?

To present our approach to these questions, this paper is organized as follows: First we introduce some background literature on how Twitter data has been used in event analysis, measuring sentiment of tweets, and a brief review of other work using social media in evaluating election outcomes. We then present our key approach to collecting data followed by the analysis. Lastly, we follow the results with discussion and next steps for projects stemming from this effort.

RELATED WORKS
Twitter Election Prediction
Political campaigns using social media has become common practice. Predicting US congressional elections based on tweets proved less efficient than predicting them using poll analyzing techniques, because knowledge of people discussing elections in Twitter are scant [1]. Although electoral prediction using tweets seems simple and straight forward issues remain. Following are additional examples of studies that have attempted to find variables that can serve as reliable predictors of election outcomes.
To investigate, authors of [2] used the Twitter API to collect data and applied many different methods for filtering and predicting. Results show that predicting elections based on tweets is too dependent on the period of data collection and too unstable because of parameterizations. Their assessment was that using tweet counting alone was not a reliable indicator for predicting elections.

Using Twitter instead of googling, before and after US presidential election 2008, authors of [3] collected the tweets in accordance with the size of each state that played a major role. Samples correctly predicted Obama would be the winner; however, the results showed a landslide victory for Obama in every state showing cases of unreliability with the model. In [4], the researchers revealed that data from social media did only slightly better than chance in predicting election results in the last US Congressional elections via reviewing the findings of other researchers and then try to duplicate their findings both in terms of data volume and sentiment analysis.

Using supervised learning and volume-based measures for sentiment analysis, authors of [5] showed that analyzing tweets can be an effective way to predict voters’ intentions. On predicting, volume of the data is biggest predictive variable followed by inter-party sentiment. Authors of [6] examined the characteristics of the three main parties in the U.K. 2010 general election, classified terms related to each, and counted the number of hashtags and URLs for each party. Experimental results showed that the proposed classification method, based on Bayesian frameworks, is capable of achieving an accuracy of 86% without any training. Collecting data through Twapperkeeper and processing with the command-line tool Gawk, [7] shows that collecting data for specific hashtags can help us track collective and individual interest and connection with certain events.

Social media takes an important role in a regulated society, but can also have unintended drawbacks depending on the government characteristics. In [8], authors examine Twitter’s role during Iran’s 2009 election crisis to understand how ordinary people use social media to gain power against the regime. They found that Iranian Twitter users did not know that the security agency can use Twitter to identify, locate and kill them. Social media like Twitter can provide information sources for protestors but not the physical power to go against the regime. With exponentially increased number of users, social media became a great tool to track people’s political preferences in Italian and French elections. Authors of [9] employed a statistically improved sentimental analysis method and demonstrates consistent correlation between tweets prediction and traditional poll survey. Accuracy improves when focusing more popular leaders or parties. Microblogs have made the online campaign possible. Authors of [10] archive tweets and SPSS to perform statistical analysis. This research suggests structural aspects of Twitter do not use the qualitative data; therefore, research needs to go beyond structural aspects to better understand Twitter use for election prediction.

Other tools geared toward language semantics have been applied to aid in the validity of user tweets. Authors of [11] collect Dutch Twitter messages for the Dutch Senate elections of 2011 with the filter stream provided by Twitter, and apply a language guesser developed to filter out irrelevant Tweets. Additionally, since there is no sentiment analysis software for the Dutch language. They constructed a corpus of political tweets with manual sentiment annotation for sentiment analysis. Despite having no gold standard training data, the total error of the final system was only 29% higher than that of two experienced private polling companies.

A Study of the 2011 Singapore General Election by [12] seeks to test validity of Twitter message on political public opinion during the 2011 Singapore General Election. Authors developed their own Twitter crawler using the Perl programming language, a MySQL database, and the application programming interface (API) provided by Twitter. They found the strong correlation between the share of Twitter messages and actual result of election. Interestingly, the ruling party received a far smaller percentage of tweets than its percentage of votes, while its opposition parties received a significantly larger percentage of tweets than the percentage of their vote.
In [13], authors used the Twitter API to collect data and identify supporters for each party by sentiment analysis, machine learning, and t-test analysis. They found that the network agendas of candidate supporters were positively correlated with the network agendas of various media types during most of the 2012 U.S. election period. In [14], authors find the difference in usage of social media during the midterm elections in the US using text mining and graph structure and find significant relationships between content, graph structure and election results by building a model that predicts whether a candidate will win or lose with accuracy of 88.0%.

In [15], authors use Twitter data on consumer confidence and presidential job approval polls over 2008-2009. Methods used include text analysis, opinion estimation, moving average aggregate sentiment, correlation analysis, and forecasting analysis. Many techniques from traditional survey methodology can be used for automatic opinion measurement, but more advanced NLP techniques can improve opinion estimation. Authors of [16] describe a system for real-time analysis of public sentiment toward presidential candidates in the 2012 U.S. election as expressed on Twitter. The real-time data processing infrastructure, based on IBM’s InfoSphere Streams platform, enables analysis and visualization modules and assembles them into a real-time processing pipeline. Their method and statistical sentiment model can evaluate public sentiment effectively.

Considering that the landscape for analyzing Twitter micro blog data is of interest by many researchers considering the cited studies above, this study is the initiation of further studies to add to the current knowledge of analyzing Twitter data in the context of political outcomes. Given the exploratory spirit of this study, we maintained a simple focus based on the U.S. Presidential Primary race of 2016 to collect data to see what can be learned by analyzing those outcomes. To guide the data collection and analysis, we emphasized our efforts to answer the following questions:

- Can election outcomes be predicted using data from Twitter account feeds?
- What are the factors that have the greatest significance toward contest outcomes?
- Does public sentiment matter?

**METHODOLOGY**

**Data Collection**

Twitter data was collected from two sources using the Twitter REST API and the Twitter STREAMING API. Collecting tweets using the Twitter REST API has limitations depending on which calling method is used. For this study, we used the GET_status method to request tweets directly from a user’s timeline. With this method, each request can only retrieve up to 200 distinct tweets, regardless if they were original from the user or if they were posted as retweets. The Twitter REST API also limits the amount of requests using the GET_statuses method to 180 requests per 15-minute window (user authorization) [17, 18]. With these limitations in mind, a scheduled call was made every 15 minutes for each of the primary candidates’ main Twitter accounts collecting their most recent tweets (Table 1).

A second source to collect Twitter feeds was established using Twitter’s STREAMING API. The STREAMING API allows requests to retrieve live tweets directly from the Twitter stream given a set of search parameters for an established amount of time. When using the STREAMING API, all live tweets are fed to a collection location over a set duration of time specified by the calling application. Due to the limitations in processing large files directly, the final strategy employed was to open the stream more frequently for short durations (e.g. every 5 minutes for 3 minute lengths). The calls to the STREAMING API was derived of search terms comprised of candidate names, candidate nicknames, and candidate handles (Table 2).
TABLE 1  
ASSOCIATE TWITTER HANDLES FOR REST API  
GET_STATUS

<table>
<thead>
<tr>
<th>Political Figure</th>
<th>Twitter Handle</th>
<th>Party Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donald Trump</td>
<td>@realDonaldTrump</td>
<td>Republican</td>
</tr>
<tr>
<td>Ted Cruz</td>
<td>@tedcruz</td>
<td>Republican</td>
</tr>
<tr>
<td>Marco Rubio</td>
<td>@marcorubio</td>
<td>Republican</td>
</tr>
<tr>
<td>Ben Carson</td>
<td>@RealBenCarson</td>
<td>Republican</td>
</tr>
<tr>
<td>John Kasich</td>
<td>@JohnKasich</td>
<td>Republican</td>
</tr>
<tr>
<td>Jeb Bush</td>
<td>@JebBush</td>
<td>Republican</td>
</tr>
<tr>
<td>Hillary Clinton</td>
<td>@HillaryClinton</td>
<td>Democrat</td>
</tr>
<tr>
<td>Bernie Sanders</td>
<td>@BernieSanders</td>
<td>Democrat</td>
</tr>
</tbody>
</table>

TABLE 2  
‘Q’ SEARCH KEYWORDS FOR STREAMING API  
(OPEN SEARCH)

<table>
<thead>
<tr>
<th>Full name</th>
<th>Single word search</th>
<th>@mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>donald trump</td>
<td>trump</td>
<td>@realdonaldtrump</td>
</tr>
<tr>
<td>ted cruz</td>
<td>tcruz</td>
<td>@tedcruz</td>
</tr>
<tr>
<td>marco rubio</td>
<td>mrubio</td>
<td>@marcorubio</td>
</tr>
<tr>
<td>ben carson</td>
<td>carson</td>
<td>@realbencarson</td>
</tr>
<tr>
<td>john kasich</td>
<td>kasich</td>
<td>@johnkasich</td>
</tr>
<tr>
<td>jeb bush</td>
<td>jbush</td>
<td>@jebbush</td>
</tr>
<tr>
<td>hillary clinton</td>
<td>clinton</td>
<td>@hillaryclinton</td>
</tr>
<tr>
<td>bernie sanders</td>
<td>sanders</td>
<td>@berniesanders</td>
</tr>
</tbody>
</table>

*single word keywords were established based on frequent references to candidates in preliminary tweet review

The overall data collection using both the REST API\(^1\) and the STREAMING API\(^2\) were set-up to operate automatically through the end of February 2016 and through the month of March 2016 to capture social media responses of major contests (e.g. Super Tuesday).

Data Analysis
The data collected from the REST API was used to determine how often candidates mention other candidates in their candidacy Twitter accounts and if the relationships provide any evidence regarding winners of election contests. All tweets from candidate user accounts were compiled into a single table in an Excel Spreadsheet. The text of all tweets were assessed to determine if the user’s discussion was directed toward themselves, another candidate in the election, or a simple non-directed rallying tweet. The resulting data was then analyzed using a market basket analysis using Excel’s Data Mining plug-in linked to Microsoft’s data mining tools.

\(^1\) Tweepy, a Python based package, was used for establishing a connection to the REST API (see http://www.tweepy.org/)

\(^2\) An R based packaged was set up for accessing the STREAMING API. For an example of the process of pulling tweets and conducting sentiment analysis, see: http://datascienceplus.com/sentiment-analysis-on-donald-trump-using-r-and-tableau/
The output of the market basket analysis resulted in illustrating that the candidates of the two major parties typically mentioned candidates that were in direct competition within their associated political party (Figure 1). Interestingly, the analysis suggested that the majority of candidates in the Republican Party pointed the majority of their comments toward the leading front runner, Donald Trump. Interestingly, Hillary Clinton, a leading candidate of the opposing Democratic Party, also had a strong association with tweeting about the opposing party’s front runner. This suggests there was some identification to respond to the opposing party candidates through the primary election process. Furthermore, it is noted that Ben Carson never had an opposing comment toward Donald Trump and subsequently over course of the data collection had dropped out of the race.

![Figure 1. Association Analysis of Candidate Mentions](image1.png)

Further analysis showed the stronger relationship to be associations between Donald Trump and Ted Cruz (Figure 2), and a relatively equally distributed discussion directed toward Ted Cruz from opposing candidates, Donald Trump (26%), Marco Rubio (23%), and Ben Carson (22%) (Figure 3).

![Figure 2. Association Analysis Item Set Support](image2.png)

![Figure 3. Association Analysis Rule Probability](image3.png)

The data collected from the STREAMING API, representing a sample of the population of tweets from the public, was used to evaluate both the user sentiment and perform a regression analysis on the factors that were significant in predicting the winners of contests over the month of March 2016. First a simple lexicon
based sentiment analysis was performed using an R script to evaluate the text [19] (for a complete example see http://analyzecore.com/2014/04/28/twitter-sentiment-analysis/) [20]. Although minimal in nature, the script evaluated each word in each tweet against a dictionary of words that were semantically either negative or positive. A composite score summing the total positive and negative words was created representing the sentiment of the tweet\(^3\) [21].

The data was assembled to perform an analysis surrounding each primary contest comprising of the data for each candidate associated with each event. To establish an observation for a single event as cross-sectional data, we assembled the data from the day before a contest, the day of the contest, and the day after the contest. This included the dependent variable as the number of delegates won per contest and the following independent variables:

- **Number of tweets**: The total number of tweets directed toward a candidate (e.g. Donald Trump) from the entire population of collected tweets.
- **Number of retweets**: The total number of tweets that were ‘retweeted’ directed toward a candidate from the entire population of collected tweets.
- **Sentiment (Sum)**: The accumulated sum of all the sentiment values directed toward a candidate
- **Average sentiment**: The accumulated sum of the sentiment over the number of tweets directed toward a candidate
- **Number of tweets by verified user account**: The total number of tweets directed toward a candidate derived only from verified users.
- **Number of retweets by verified user account**: The total number of tweets that were ‘retweeted’ directed toward a candidate derived only from verified users.
- **Sentiment from verified users (Sum)**: The accumulated sum of all the sentiment values directed toward a candidate derived only from verified users.
- **Average sentiment from verified user accounts**: The accumulated sum of the sentiment over the number of tweets directed toward a candidate derived only from verified users.

Using SPSS, a linear regression was performed on all the data to determine which factors, if any, were significant in predicting the winning count of delegates of each contest. The overall descriptive statistics are summarized in Table 3 and the results of the regression analysis are portrayed in Table 4.

\(^3\) This method is a simple semantic analysis comparing each word in the tweet against a lexicon library of words that have been tested to be perceived as either positive or negative in valence. A complete example of R based code can be found on this website: http://analyzecore.com/2014/04/28/twitter-sentiment-analysis/ . The key components are comprised of the lexicon library for comparison. For this study, a lexicon of approximately 6800 words were used available at this site: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon. The process entails each word in each tweet to be compared to the corpus of words in the lexicon to determine if it is positive or negative. A sum value of positive and negative words is calculated for each tweet, then the semantic score of the tweet is calculated by summing the count of positive words less the count of negative words. This method has received some criticism due to the potential for double negative in phrases, however given nature of tweets, it is felt that this method can provide an indicator toward the exploratory investigation of sentiments in user tweets.
TABLE 3
DESCRIPTIVE STATISTICS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SE</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delegates</td>
<td>0</td>
<td>522</td>
<td>76.4</td>
<td>22.2</td>
<td>122.</td>
</tr>
<tr>
<td>No of Tweets</td>
<td>1,259</td>
<td>2,885,578</td>
<td>404,135</td>
<td>117,885</td>
<td>645,685</td>
</tr>
<tr>
<td>No of Retweets</td>
<td>54,926</td>
<td>4,517,777,55</td>
<td>405,340,208</td>
<td>175,109,249</td>
<td>959,112,860</td>
</tr>
<tr>
<td>Sentiment</td>
<td>-35,970</td>
<td>227,480</td>
<td>19,702</td>
<td>8,374</td>
<td>45,868</td>
</tr>
<tr>
<td>Avg Sentiment</td>
<td>-0.0242</td>
<td>0.167</td>
<td>0.047</td>
<td>0.0086</td>
<td>0.045</td>
</tr>
<tr>
<td>Ver No of Tweets</td>
<td>27</td>
<td>23,908</td>
<td>3,777</td>
<td>991</td>
<td>5,428</td>
</tr>
<tr>
<td>Ver No of Retweets</td>
<td>42</td>
<td>1,866,478</td>
<td>262,267</td>
<td>87,821</td>
<td>481,016</td>
</tr>
<tr>
<td>Ver Sentiment</td>
<td>-64</td>
<td>3,390</td>
<td>805</td>
<td>213</td>
<td>1,170</td>
</tr>
<tr>
<td>Ver Avg Sentiment</td>
<td>-0.1429</td>
<td>0.566</td>
<td>0.221</td>
<td>0.036</td>
<td>0.195</td>
</tr>
</tbody>
</table>

Note: N = 30

Given the data set was derived from a panel of time related data, one potential threat to the validity of the data is the issue of autocorrelation and the effects that previous events had on following event outcomes. Autocorrelation is common threat to statistical validity when data is analyzed as cross-sectional, yet derived from serial or time based information and having the potential to carry effects from one measurement onto another. To alleviate this threat is to test for autocorrelation by calculating a Durbin-Watson statistic. In general, a Durbin-Watson statistic is a test statistic used to determine if the threat of autocorrelation exists within data when conducting a regression analysis. For this analysis, the Durbin-Watson statistic was calculated with SPSS resulting in a value of 2.79, which is above the preferred limit of 2\[22, 23\]. With this outcome, there were no concerns to perform additional transformations to the data for analysis.

The regression analysis showed that the factor that was significantly supported in predicting the number of won delegates is the sentiment from verified users (\(\rho = 0.003, SE = 0.040; R^2 = 0.88\)). As an effort to prevent false brands and identities representing key individuals, Twitter has a process that verifies user accounts to ensure those posting tweets are who they say they are. The verification process is rigorous and Twitter does not allow verified user accounts from general public, distinguishing authentic accounts from fan-based or non-associated users of popular trademarks. The findings from the regression analysis suggest that the sentiment generated from verified user accounts are significant predictors in the outcomes of a primary contest. This may be because verified users are sound boards, echoing the feelings and sentiment of public preferences, or they could also be a voice that influencing public preference. However, with this study, the directionality of influence is unable to be determined without further investigation. Other factors within the regression model failed to provide support for contest outcomes.
### TABLE 4

**REGRESSION OUTPUT – DEPENDENT VARIABLE NUMBER OF DELEGATES WON**

<table>
<thead>
<tr>
<th>Variable</th>
<th>B Coeff</th>
<th>SE</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of Tweets</td>
<td>0.000</td>
<td>0.000</td>
<td>0.476</td>
</tr>
<tr>
<td>No of Retweets</td>
<td>0.000</td>
<td>0.000</td>
<td>0.116</td>
</tr>
<tr>
<td>Sentiment</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.194</td>
</tr>
<tr>
<td>Avg Sentiment</td>
<td>555.9</td>
<td>336.2</td>
<td>0.144</td>
</tr>
<tr>
<td>Ver No of Tweets</td>
<td>-0.009</td>
<td>0.019</td>
<td>0.635</td>
</tr>
<tr>
<td>Ver No of Retweets</td>
<td>0.000</td>
<td>0.000</td>
<td>0.537</td>
</tr>
<tr>
<td>Ver Sentiment</td>
<td>0.138</td>
<td>0.040</td>
<td>0.003**</td>
</tr>
<tr>
<td>Ver Avg Sentiment</td>
<td>3.931</td>
<td>95.975</td>
<td>0.968</td>
</tr>
</tbody>
</table>

*Note:* **p < .05 level, * p < .10 level, N = 30, R^2 = 0.882**

### CONCLUSIONS AND LIMITATIONS

As the nature of this study being exploratory to investigate the options that can be garnered from Twitter data during an election, our goal was to initiate an effort into the use of social media outlets in collecting predict public views and how they related to political contests. Within this context, our intention is to explore how candidate participation within elections correspond to the public views and the political landscape of the election. Conducting association analysis using Twitter accounts associated with competing candidates, we observed how relationships were perceived with respect to be the leading candidate (Donald Trump) and how that influenced the frequency of mentions in Twitter account feeds. Furthermore, we witnessed how the conversation created between Donald Trump and Ted Cruz exemplified a prospective rivalry demonstrating a struggle between first and second position in the race. Further findings show that there were targeted mentions between Clinton and Trump regardless of the fact they were competing in different political races. A last interesting note were the lack of mentions of the early Republican front runner Ben Carson, who later dropped out the race.

By nature of this study, it is limited in these association observations by not statistically exploring how these relationships are directed in valance (e.g. sentiment) and in time variance. Also, it is unclear if the extended mentions of the Republican front runner from other candidates also contributed to his growth in popularity in the race. This would be a good opportune area for future research to understand the nature of the conversations and their effect on candidate popularity and election outcomes. This can be expanded to include the discussions through news outlets and other media.

Another goal of this exploration was to evaluate if social media can be a predictor of contest winners. Based on the linear regression event analysis, we can see there is some validity toward verified user sentiment as a predictor toward election contests. This finding suggests there is some relationship with verified Twitter users and public political views. Verified users are deemed by Twitter as being popular figures in society which can explain their potential to have influence on public views. However, it is not feasible to assess the direction of causality of this relationship if public opinion also influence the views of the verified users.

One limitation in evaluating the predictors of election outcomes for this study is the relatively small sample size generated from the events of a single month of the presidential primary contests. With extended data...
collection, it possible to confirm not only if verified user sentiment is robust predictor of voter results, but expanding the data collection may provide the variation required for variables smaller in effect size to be recognized as significant. The findings of this study does provide a direction for future research toward the influence of verified user accounts in social media to determine if they are a reflection of voter preferences, or if they establish them.

One question we were not able to address is if the general public sentiment, either positive or negative, really had an impact on election results. With the general user sentiment not being significant in the regression model, it is difficult to show support in how the public sentiment influences voter responses. One way to potentially address this question would be to evaluate public sentiment over multiple social media platforms and explore if the combined tone of user opinions have an impact.

In summary, the exploratory study provided some insight into the U.S. Presidential primary elections over the spring of 2016. With that, we were able to show some research that can be conducted investigating political contests using social media user feeds. With the study, future ideas can be explored, such as how verified user sentiment has an impact on general user perceptions related to public social issues involving businesses, public health, or government activities.
REFERENCES