International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 06, Volume 07 (June 2020) https://www.irjcs.com/archives

BIG DATA & SENTIMENT ANALYSIS USING PYTHON

Akshansh Malik
UG Scholar, Department of CSE, IMS Engineering College, Ghaziabad, UP, India
Akshay.malik659@gmail.com;
Aman Mahal,
UG Scholar, Department of CSE, IMS Engineering College, Ghaziabad, UP, India
mahaalaman@gmail.com
Aman Kamboj
UG Scholar, Department of CSE, IMS Engineering College, Ghaziabad, UP, India
Amankamboj7768@gmail.com
Abhishek Sharma
UG Scholar, Department of CSE, IMS Engineering College, Ghaziabad, UP, India
Abhisheksharma07041999@gmail.com

Publication History
Manuscript Reference No: IRJCS/RS/Vol.07/Issue06/JNCS10089
Received: 05, June 2020
Accepted: 16, June 2020
Published:26, June 2020
DOI: https://doi.org/10.26562/irjcs.2020.v0706.003
Citation: Akshansh,Aman,Abhhek & Aman(2020). Bigdata & Sentiment Analysis Using Python. IRJCS:: International Research Journal of Computer Science, Volume VII, 159-166.**DOI**: https://doi.org/10.26562/irjcs.2020.v0706.003
Peer-review: Double-blind Peer-reviewed
Editor: Dr.A.Arul Lawrence Selvakumar, Chief Editor, IRJCS, AM Publications, India

Abstract: "In present time, social media is having a lot of information of its users. Some of the ways are to extract information from social media gives us several usage in various types of fields and researches. In Product Analysis, extracting information from social sites/media is providing number of advantages such as knowledge about the latest technology, update of a real-time situation in market etc. one of the social media is Twitter which allows the user post tweets of limited number of characters and share the message(tweet) to their followers. It allows developer to access the information for their purpose. In the implemented module, details collected and sentiment analysis is performed on it. Based on the results of data & sentimental analysis tips and information can be provided to the user. The running module can perform data &sentiment analysis on data available for various fields and consumer opinions and suggestion on various products. These results can provide to companies to get up-to-date, etc. With this process, the implemented system can help in predicting the effects of various products and various activities in various fields."

Keywords: Big ; data; Sentiment; analysis; Python;

I. INTRODUCTION

In the present era, Social media and many corresponding applications allow all its users to express their opinions about a particular topic and show their attitudes by liking or disliking content. All are continuously accumulating actions on social media and generating high variety, volume, velocity, value, variability data termed as big social data. This kind of data refers to massive set of opinions of individual that can be processed to understand the people tendencies in the digital world. Various researchers have shown a keen interest in the exploitation of huge social data in order to explain, determine and predict human mindset in several domains. To Process this kind involve various research avenues, particularly, text analysis. In fact, 85% of online data is text, and analysis of text data has become key element for finding the sentiments of public and their valuable opinion towards the content. Sentiment analysis is also called opinion mining, which aims to find out the sentiments of users about a topic by doing analyses of their posts and different types of actions on social media. Then, it polarity is going to be classified into three categories such as positive, negative and so on.

Sentiment analysis can be divided into two categories:

- Lexicon analysis, which aims to find or calculate the polarity of a document from the semantic analysis of words or phrases in the document. However, applications which are based on lexicon analysis never consider the studied context.
- Machine learning (ML), it involves formation of models from labeled training dataset (instances of texts or sentences) in order to determine the orientation of words and phrases in a document. Studies that used machine learning methods have been carried out on an important topic.

These above two analysis methods have been widely used on big data to gather public critics in order to assess internaut's satisfaction of a subject (services, products, events, topics or different persons) in different domains including health, politics and marketing. However, the results sometimes can be varying with a reasonable degree of accuracy and sometimes are not. The failure is generally arises due to the challenges of opinion mining such as the semantic analysis of a word whose meaning depends upon the context. In this paper, our aim is to tackle semantic analysis by introducing an efficient and novel adaptable approach that depends on social media posts and architecture of big data to analyze internaut's feelings and behavior toward a particular subject in real-time. The proposed approach is based on three stages as shown in Fig. below.

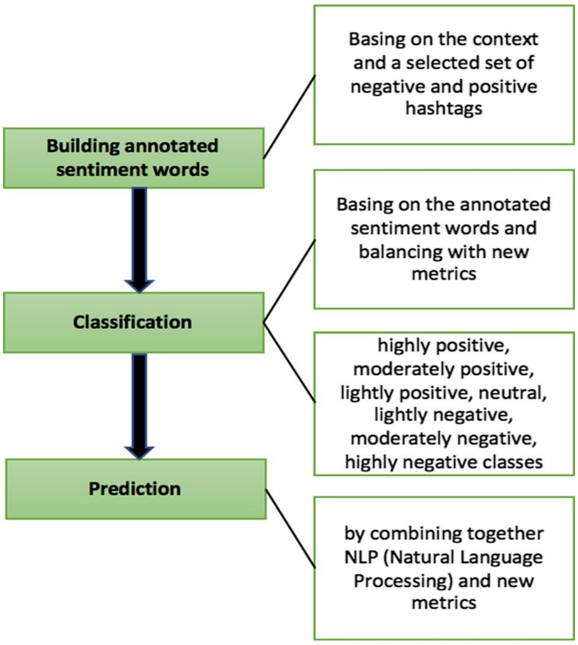


Fig 1. Stages of the model

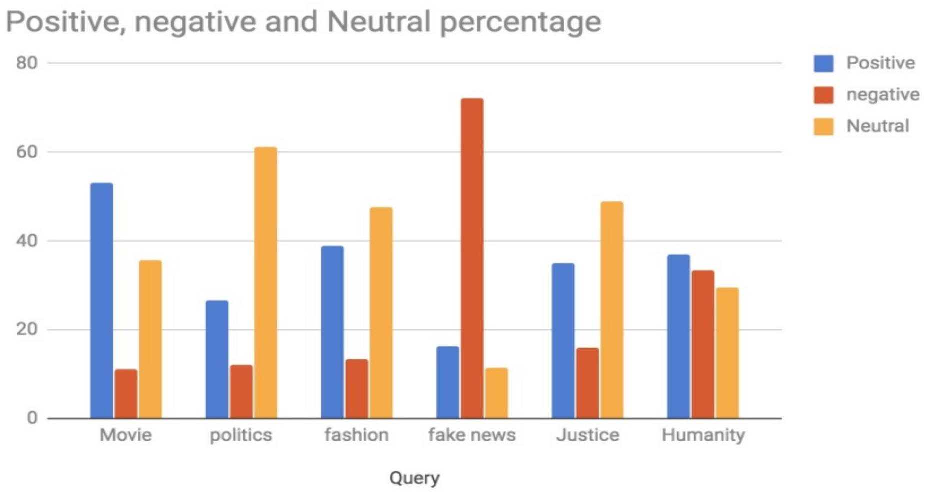
II. RELATED WORK

Several studies have focused on analyses of social media especially when it is related to some big events that are going to have a great attention like presidential elections. Social media became an important platform for both, candidates to get in direct touch with people and to share their programs, and voters to express their views about each candidate. This intensive use of social media platforms has attracted wide attention in academic research and many types of contributions have been conducted to follow that kind of events which includes millions of internet or social media users.

These contributions could be categorized into four categories:

- Opinion-based approach in which opinion mining methods are used by the authors for their models, detailed in "Theoretical basis" section, in order to classify posts related to a candidate. There are two categories classes:
  - Sentiment classes, in which sentiments are, classified under two classes (positive and negative). Other authors have added a third class (neutral).
  - Context classes, in which polarity depends on the context. Conover et al. have classified the posts in three classes which are Left, Ambiguous and Right while authors in have based their prediction model on two classes (Pros and Anti) for each party.
- Volume-based approach, in this researchers aim is to determine the candidate who is elected based on the number of tweets which mentioning them (% mention) or retweet volume. In fact, researchers in have discovered an interesting correlation between name mentions percentage or retweet volume and the vote.

- Opinion and volume (OV)-based approach in which both opinion mining and volume approaches are combined together. Researcher Finn et al. discovered a new approach to measure political polarization without the use of text. They have used a co-retweeted network, as well as the retweeting behavior of the social media users.
- Emoji-based approach, in this the classification of posts is based on the use of emoji. Researchers selected differente moji and categorized them into different categories such as happy, sad, fear, laughter, and angry classes, then, find the sentiment of the first emoji in the post.



2.1 Text mining

Text mining can be stated as a process of determining useful information from textual data which is present in unstructured manner. This process is based on two major phases: The analysis phase which refers to the process of structuring text by using linguistic analysis techniques such as recognizing the words, sentences, their grammatical roles and relationships. It involves several methods:

- Language identification, it is the process of determining the natural language in which a given text is written.
- Tokenization, it is the process of segmenting a sequence of strings into words and sentences by removing some characters like punctuation marks.
- Filtering, this process consists of applying filters such as removing empty words.
- Lemmatizations in this process different inflected form of a word are grouping together by removing plurals, genders and conjugations. Then, we analyze them as a single item.
- Named-entity recognition, it is the process in which we are going for searching text object which can be categorized in classes such as persons, dates and localization.

The output of the first phase is evaluated and interpreted by interpretation phase by using methods of data mining. The purpose is to find patterns, relevance, novelty, and interestingness. Text mining does not allow extracting opinion, so other techniques are also combined in it. Next, we will present some of those techniques.

2.2 Opinion mining

Opinion mining is the technique of science in which we are using text analysis to determine the sentiment analysis of a text (positive, negative or neutral). It can be determine under different terms: sentiment analysis subjectivity, analysis of stance. One of its important application is to understand and track the mood of the users of social media about a specific topic in different domains such as marketing, health and politic. These product reviews can be used by Potential buyers for making their decisions related to the product.

These are following approaches to carry out opining mining:

2.3 Lexicon based approach

Sentiment lexicon and a collection of known sentiment terms are used in lexicon-based approach. It is mainly divided into two approaches i.e. dictionary-based approach and corpus-based approach. Dictionary-based approach finds opinion words in the text and then finds semantic orientation of those words in the dictionary.

There are number of dictionaries such as Senti Word Net and they can also be created manually. Corpus-based approach is used to find opinion words in a context specific orientation. It generally starts with a list of opinion words and then other opinion words are going to be determined in a large corpus.

2.4 Learn based approach

The learn-based approach depends on the famous ML algorithms (i.e. supervised and unsupervised methods). In the supervised methods, we will train the model through a large number of labeled documents. The most famous ones for opinion mining are: Support vector machine, maximal entropy principle, and the Naive Bayesian classification. These methods having a high accuracy in determining the polarity in the domain that they are trained on but their performance fall precipitously when the same model is used in a different domain. The unsupervised methods are used when it is difficult to find labeled training documents. However, because of their bad performances in this area, it is rarely used at present.

2.5 Hybrid approach

The hybrid approach uses both, the lexicon and the learn-based approaches. It uses the lexicon-based approach for determining the sentiment scoring. Then, training data for the learn-based part will represent by these scored documents. Hybrid approach is widely used because of its two qualities i.e. improved or high accuracy and due to its stability that comes from ML powerful and the lexicon based approach, respectively.

All three approaches i.e. Hybrid, learn-based and Lexicon-based approaches have been widely used in different domains, in different ways and improved their efficiency by several searchers.

By referring to opinion mining approaches, we present a method that analyzes social media posts and extracts user's opinion.

III.THE PROPOSED METHOD

In order to build a model for sentiment analysis, we propose three stages based methodology in this paper that contains, first building sentiment words, then classifying and balancing this set of words before executing the prediction algorithm.

The descriptions of the three stages are explained below:

Let: Y_i, $i=1,n$ be a set of products, services or persons that we will aim to compare in a specific context.

Let's consider $D= \{Y_1, Y_2, \dots, Y_n\}$.

3.1 FIRST STAGE: CONSTRUCTING DICTIONARIES

Based on hashtags description various researches in social media analysis have identified whether the intention behind a post is positive, negative or neutral. However, they have used a manually defined very large set of annotated hashtags (which may take large amount of time) or they have combined these latter with dictionaries in order to improve classification posts accuracy. At the beginning stage, we use a small set of hashtags, in order to build dictionaries of words, annotated with the word's semantic orientation for a given context as following:

We are going to assume that each and every word in a tweet that contains a negative hashtag is negative and a tweet that contains a positive hashtag is positive, then, we process it by different steps.

- **Step1** Contains posts which are related to Y_i. The aim of our approach is to compare Y_i, we will going to identify hashtags with high frequency as the most popular hashtags for each Y_i. After that we will classify a small set of them manually into negative and positive classes and have to collect related data for each class separately. We will classify collected tweets into two classes i.e. positive or negative based on the upper defined polarity of hashtags.
- **Step 2** Consists of data which is preprocessing classified extracted from hashtags. Social data is an informal type of data that could contain spelling mistakes and non-textual information, hence there is need of a preprocessing step and for that we are going to apply various filters on tweets as following:
  - Tokenization is a process that contains sub-step which consists of identifying nouns, verbs, adverbs, adjectives, URLs, common emoticons, phone numbers, HTML tags; Twitter mentions hashtags, and repetition of symbols and Unicode characters.
  - Conversion is a process where all words will be converted to lowercase and replace more than two of the same consecutive letters in a word with only one occurrence of the letter (e.g., we replace Sunny by sunny and ANGER by anger).
  - Stemming is a process where we remove plurals genders and conjugation (applying morphology stemming).
  - Filtering can be defined a process where we enhance the indicators by applying other various filters and sentiment indicators such as hashtags.
  - Indicators can be defined as adjectives and verbs which are good indicators for positive and negative sentiment analysis. However, as social data could contain more information than a formal text, we The output of this step will represent the intermediate sentiment words of each: interpose SW(Y_i) and inter-neg SW(Y_i)

- **Step 3** The purpose of this step can be defined as to refine the annotated dictionary: positive posSW(), negative negSW()and neutral neutSW() dictionaries for each Y_i . The task of classification of neutral hashtags is difficult and that could affect the result, so that's why we ignored them during the collect. In fact, a tweet that contains a neutral hashtag such as #modi could be either negative or positive. Therefore, we have to construct neutral basing on the word occurrence $Occ(w_j)$ for all in the different classes. This will allows us to construct the final dictionaries by using of Algorithm 1.

We conducted empirical test that consists of testing a number of values (between 0.5 and 0.8), in order to constitute the limit that allows classifying sentiment words with the smallest error rate. In our case 0.7 was the best value. At the end, we will assign a score to sentiment words: 1, 0, − 1 for positive, neutral and negative, respectively.

Algorithm 1 Extract Twitter sentiment

```
procedure TWITTER-CONNECTION()
    consumer − key =' xxxxxxxxx'
    consumer − secret =' xxxxxxxx'
    access − token =' xxxxxxxx'
    access − token − secret =' xxxxxxxxxx'
    self.auth = OAuthHandler(consumer − key, consumer − secret)
    self.auth.set − access − token(access − token, access − token − secret)
    self.api = tweepy.API(self.auth)
end procedure

procedure TWEET-CLEANING(t)
    tweet = t.remove − Stop − words
    Return tweet
end procedure

procedure TWEET-CLASSIFICATION(t)
    t = Tweet − Cleaning(t)
    tweet − polarity = t.sentiment.polarity
    tweet − polarity
end procedure

procedure GET-TWEETS(q, count)
    fetched − tweets = self.api.search(q = query, count = count)
    Return fetched − tweets
end procedure

procedure MAIN()
    st = SentimentalTwitter()
    tweets = st.fetch − tweets(query =' politics', count = 300)
    PositiveTweets = tweetsthatsentiment =' positive'
    NegativeTweets = tweetsthatsentiment =' negative'

    for tweet t in PositiveTweets do
        print(t)
    end for
    for tweet t in NegativeTweets do
        print(t)
    end for
end procedure
```

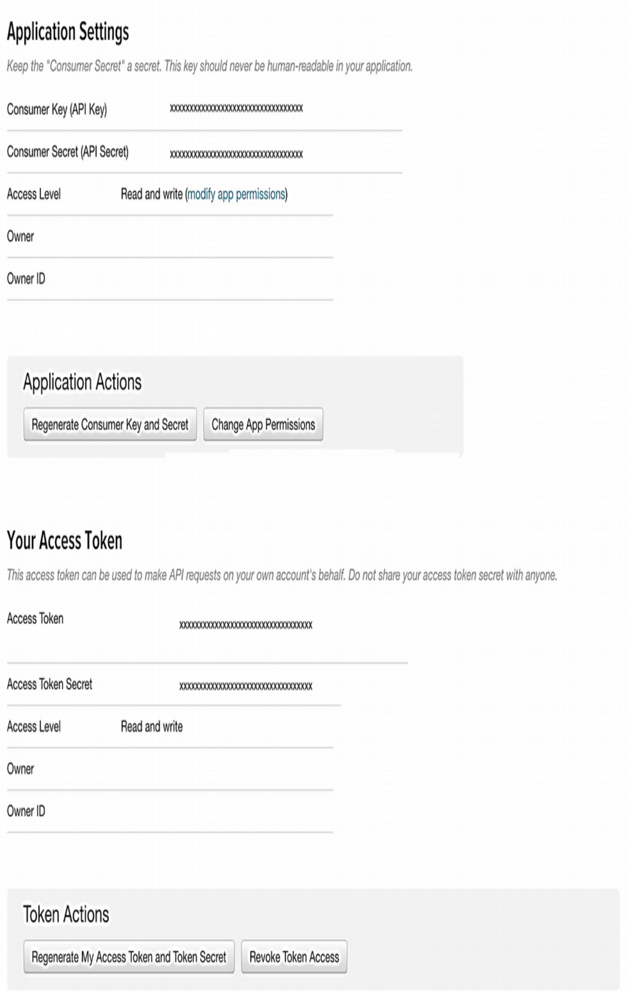


Figure 2 shows the modules of the above shown first stage:

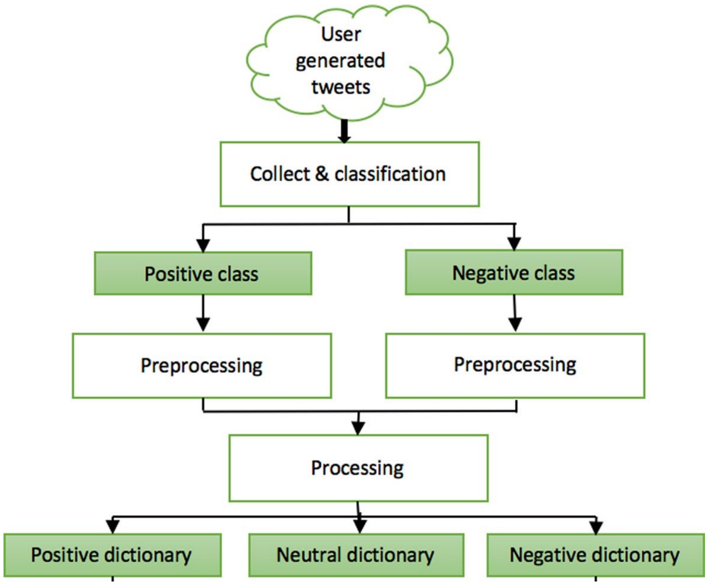


Fig.2 First stage of the model

3.2 SECOND STAGE: CLASSIFICATION

In this stage, we will classify new tweets based on the *SW* dictionary build in the previous stage.

- **Step 1** Includes the process of collecting and storing new tweets for each Y_i separately. Collected data goes through the following number of steps.
- **Step 2** Includes the process of preprocessing the data as following:
  - Removing duplicated data can be defined as a process of removing duplicated tweets in order to avoid misleading results. As a post could include multiple hashtag and it could be extracted for multiple times.
  - Tokenization is the same process that we used in the first stage is applied to the new used tweets.
  - Handling negation, Here negation words can be defined as words like (no, not, nothing and nonce). These words can significantly affect the overall polarity of a sentence, it is considered as a very important criterion in the field of sentiment classification. As stated in researches we reverse the sentiment polarity of the words that come after a negation word until reaching a punctuation mark.
  - Handling repetition is a process in which words are going to be detected that are written in uppercase or constitute more than two of the same consecutive letters.
  - Applying morphology by following the same rules as the first stage.
- **Step 3** In this step, we are going to calculate degree of polarity of the tweets based on their semantic orientation of words which are assigned in stage 1. For this, we are going to apply the following two actions:
  - Balancing It is worth noting that the used language in social media posts is not conventional and could contain some special words such as those that are written in upper case or contain the repetition of more than two consecutive letter which we call "extended word".
  - Calculating polarity degree, in this we going to find the polarity of a tweet which is calculated by adding up the independent score values of words stated in t. m is the length of t.

$$_p(t) = \sum_{k=1}^{m} score(w)$$

- **Step 4** in order to classify tweets, we use p(t) as following: We classify scored tweets into seven classes according to the polarity degree into C_{+3}, C_{+2}, C_{+1}, C_0, C_{-1}, C_{-2}, $C_{-3}$and these classes will symbolize tweets as highly positive, moderately positive, lightly positive, neutral, lightly negative, moderately negative, highly negative classes, respectively). For this, we conduct empirical test to determine the limit of each class. If case1: $0 < p(t) \leq 3$, then the tweet is classified as lightly positive. If case2: $4 \leq p(t) \leq 6$, it is moderately positive. If case3: $p(t) \geq 7$, then, it is highly positive. If case4: $-3 \leq p(t) < 0$, then, the tweet is classified as lightly negative. If case5: $-6 \leq p(t) \leq -4$, it is moderately negative. If case6: $p(t) \leq -7$, then, it is highly negative. There is also a possibility for a sentiment score to be equal to 0, if it is equal to zero, then, the tweet is classified as neutral.

3.3 THIRD STAGE: PREDICTION

Several researchers have considered three classes' i.e. positive, negative and neutral classes to determine the sentiment of a document based on the words and/or emoticons and only few ones, such Khatua et.al.

Have examined the polarity degree (i.e. highly, moderately, weakly positive and negative classes). But authors have considered only two indicators i.e. strongly positive and strongly negative.

IV. CLASSIFICATION ACCURACY EVALUATION

To assess the ability of classifying tweets based on the automatically constructing dynamic dictionary, we have randomly selected a subset of 210 tweets from the political Twitter corpora: 30 for each class. The tweets were manually inspected and labeled into classes as positive, moderately positive, highly positive, lightly negative, moderately negative, strongly negative or neutral for each candidate. Then, the same data was processed, as mentioned above, following various steps such as "by removing stop words, applying tokenization, stemming and various filters". The above step was done by the help of TreeTagger, which is a tool for annotating text with part-of-speech and lemma information. TreeTagger was also modified to handle various other things such as negation, URLs, usernames, Twitter mentions and hashtags and intensifiers.

V. CONCLUSION AND DISCUSSION

Sentiment analysis has been proven to be effective in predicting people reaction or opinion by analyzing big social data on a particular topic. The technique which we proposed consists of various steps starting with building a dictionary of words' polarity based on a very small set of positive and negative hashtags related to a particular given subject, then, posts will be classified into several classes and balancing the sentiment weight by using new metrics such as uppercase words and the repetition of more than two consecutive letter in a word.

However, the proposed approach still suffer have some challenges. First, it cannot understand emoticons. Second, we used only Twitter data. Third, we cannot access large data for this algorithm. For further improvement, we wish to handle these three limitations by proposing a more efficient and global model that can work on larger volumes of data

REFERENCES

1. Balasubramanyan R, Routledge BR, Smith NA. From tweets to polls: linking text sentiment to public opinion time series. Icwsm. 2010; 11:1–2.
2. Benamara F, Cesarano C, Picariello A, Recupero DR, Subrahmanian VS. Sentiment analysis: adjectives and adverbs are better than adjectives alone. In: Proceedings of ICWSM conference. 2007.
3. Bermingham A, Smeaton A. On using Twitter to monitor political sentiment and predict election results. In: Proceedings of the workshop on sentiment analysis where AI meets psychology. 2011.
4. Bhatt R, Chaoji V, Parekh R. Predicting product adoption in large-scale social networks. In: Proceedings of the 19th ACM international conference on Information and knowledge management. New York: ACM; 2010. p. 1039–48.
5. Chesley P, Vincent B, Xu L, Srihari RK. Using verbs and adjectives to automatically classify blog sentiment. In: AAAI symposium on computational approaches to analyzing weblogs (AAAI-CAAW). 2006. p. 27–9.
6. Conover MD, Goncalves B, Ratkiewicz J, Flammini A, Menczer F. Predicting the political alignment of twitter users. In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. 2011. p. 192–9.
7. De Choudhury M. Predicting depression via social media. ICWSM. 2013; 13:1.
8. Delenn C, Jessica Z, Zappone A. Analyzing Twitter sentiment of the 2016 presidential candidates. Stanford: Stanford University; 2016.
9. DiGrazia J, McKelvey K, Bollen J, Rojas F. More tweets, more votes: social media as a quantitative indicator of political behavior. PLOS ONE. 2013; 8(11):e79449.
10. Ekaterina O, Jukka TO, Hannu K. Conceptualizing big social data. J Big Data. 2017;4:3.
11. Finn S, Mustafaraj E, Metaxas PT. The co-retweeted network and its applications for measuring the perceived political polarization. Faculty Research and Scholarship. 2014.
12. Gayo-Avello D. No, you cannot predict elections with Twitter. IEEE Internet Comput. 2012;16(6):91–4.
13. Hansen LK, Arvidsson A, Nielsen FA, Colleoni E, Etter M. Good friends, bad news-affect and virality in twitter. In: Future information technology, communications in computer and information science. Berlin: Springer; 2011. p. 34–43. https://doi.org/10.1007/978-3-642-22309-9_5.
14. Hu M, Liu B. Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, KDD'04. New York: ACM; 2004. p. 168–77.
15. Jahanbakhsh K, Moon Y. The predictive power of social media: on the predictability of US presidential elections using Twitter. https://arXiv:1407.0622 [physics]. 2014.
16. Jose R, Chooralil VS. Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach. In: 2016 international conference on data mining and advanced computing (SAPIENCE). 2016. p. 64–7.
17. Khatua A, Khatua A, Ghosh K, Chaki N. Can #Twitter_trends predict election results? Evidence from 2014 Indian general election. In: 2015 48th Hawaii international conference on system sciences. 2015. p. 1676–85.

18. Livne A, Simmons M, Adar E, Adamic L. The party is over here: structure and content in the 2010 election. In: Fifth International AAAI conference on weblogs and social media. 2011.
19. Mahmood T, Iqbal T, Amin F, Lohanna W, Mustafa A. Mining Twitter big data to predict 2013 Pakistan election winner. In: INMIC. 2013. p. 49–54.
20. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J. 2014;5(4):1093–113.
21. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques .In: Proceedings of the ACL-02 conference on empirical methods in natural language processing, vol. 10. Stroudsburg: EMNLP'02, Association for Computational Linguistics; 2002. p. 79–86.
22. Pääkkönen P. Feasibility analysis of AsterixDB and Spark streaming with Cassandra for stream-based processing. J Big Data. 2016;3:6. https://doi.org/10.1186/s40537-016-0041-8.
23. Ramanathan V, Meyyappan T. Survey of text mining. In: International conference on technology and business and management. 2013. p. 508–14.
24. Ramteke J, Shah S, Godhia D, Shaikh A. Election result prediction using Twitter sentiment analysis. In: 2016 international conference on inventive computation technologies (ICICT), vol. 1. 2016. p. 1–5.
25. Razzaq MA, Qamar AM, Bilal HSM. Prediction and analysis of Pakistan election 2013 based on sentiment analysis. In: 2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014). 2014. p. 700–3.
26. Ruths D, Pfeffer J. Social media for large studies of behavior. Science. 2014;346(6213):1063–4.
27. Shi L, Agarwal N, Agrawal A, Garg R, Spoelstra J. Predicting US primary elections with Twitter. Stanford: Stanford University; 2012.ć
28. Smailovi J, Kranjc J, Grčar M, Žnidaršič M, Mozetič I. Monitoring the Twitter sentiment during the Bulgarian elections. In: 2015 IEEE international conference on data science and advanced analytics (DSAA). 2015. p. 1–10.
29. Soler JM, Cuartero F, Roblizo M. Twitter as a tool for predicting elections results. In: 2012 IEEE/ACM international conference on advances in social networks analysis and mining. 2012. p. 1194–200.
30. Speriosu M, Sudan N, Upadhyay S, Baldridge J. Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the first workshop on unsupervised learning in NLP, EMNLP'11. Stroudsburg: Association for Computational Linguistics. p. 53–63.
31. Stavrianou A, Brun C, Silander T, Roux C. NLP-based feature extraction for automated tweet classification. In: Proceedings of the 1st international conference on interactions between data mining and natural language processing, vol. 1202, DMNLP'14. Aachen: CEUR-WS.org; 2011. p. 145–146.
32. Tumasjan A. Predicting elections with Twitter: what 140 characters reveal about political sentiment. In: Fourth international AAAI conference on weblogs and social media. 2010.
33. Tumitan D, Becker K. Sentiment-based features for predicting election polls: a case study on the Brazilian scenario. In: 2014 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies (IAT), vol. 2. 2014. p. 126–33.
34. Tunggawan E, Soelistio YE. And the winner is...: Bayesian Twitter-based prediction on 2016 US presidential election. https://arXiv:1611.00440 [cs]. 2016.
35. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S. A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle. In: Proceedings of the ACL 2012 system demonstrations, ACL'12. Stroudsburg: Association for Computational Linguistics; 2012. p. 115–20.
36. Wang H, Castanon JA. Sentiment expression via emoticons on social media. In: 2015 IEEE international conference on Big Data (Big Data). 2015. p. 2404–8.
37. Wicaksono AJ, Suyoto P. A proposed method for predicting US presidential election by analyzing sentiment in social media. In: 2016 2nd international conference on science in information technology (ICSITech). 2016. p. 276–80.

38. Wong FMF, Tan CW, Sen S, Chiang M. Quantifying political leaning from tweets, retweets, and retweeters. IEEE Trans Knowl Data Eng. 2016; 28(8):2158–72.
39. Xie Z, Liu G, Wu J, Wang L, Liu C. Wisdom of fusion: prediction of 2016 Taiwan election with heterogeneous big data. In: 2016 13th international conference on service systems and service management (ICSSSM). 2016. p. 1–6.
40. Xing F, Justin ZP. Sentiment analysis using product review data. J Big Data. 2015; 2:5.
41. Yu H, Hatzivassiloglou V. towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 conference on empirical methods in natural language processing, EMNLP'03. Stroudsburg: Association for Computational Linguistics; 2003. p. 129–36.

