

## Analysing and Visualizing Tweets for U.S. President Popularity

Ernesto De Luca<sup>#+</sup>, Francesca Fallucchi<sup>#+</sup>, Romeo Giuliano<sup>#</sup>, Giuseppe Incarnato<sup>#</sup>, Franco Mazzenga<sup>\*</sup>

<sup>#</sup> *Department of Innovation and Information Engineering, Guglielmo Marconi University, via Plinio 44, Rome, 00193, Italy  
E-mail: ew.deluca@unimarconi.it; f.fallucchi@unimarconi.it; r.giuliano@unimarconi.it; giuseppe.incarnato@gmail.com*

<sup>+</sup> *Digital Information and Research Infrastructures, Georg Eckert Institute Braunschweig, Germany  
E-mail: fallucchi@gei.de; deluca@gei.de*

<sup>\*</sup> *Department of Enterprise Engineering “Mario Lucertini”, University of Rome Tor Vergata, via del politecnico 1, Rome, Italy.  
E-mail: mazzenga@ing.uniroma2.it*

---

**Abstract**—In our society we are continually invested by a stream of information (opinions, preferences, comments, etc.). This shows how Twitter users react to news or events that they attend or take part in real time and with interest. In this context it becomes essential to have the appropriate tools in order to be able to analyze and extract data and information hidden in their large number of tweets. Social networks are a source of information with no rivals in terms of amount and variety of information that can be extracted from them. We propose an approach to analyze, with the help of automated tools, comments and opinions taken from social media in a real time environment. We developed a software system in R based on the Bayesian approach for text categorization. We aim of identifying sentiments expressed by the tweets posted on the Twitter social platform. The analysis of sentiment spread on social networks allows to identify the free thoughts, expressed authentically. In particular, we analyze the sentiments related to U.S President popularity by also visualizing tweets on a map. This allows to make an additional analysis of the real time reactions of people by associating the reaction of the single person who posted the tweet to his real time position in Unites States. In particular, we provide a visualization based on the geographical analysis of the sentiments of the users who posted the tweets.

**Keywords**— Twitter; Russiagate; sentiment analysis; U.S. President popularity.

---

### I. INTRODUCTION

The advent of the Internet, the large amount of available data and the next document digitalization brought people to have an excessive quantity of daily information. Some problems raised in terms of storage, preservation, communication, compression, and extraction of information. It is easy to understand the content (or their opinions) of a few documents. Nevertheless, the text understanding is not an easy task when the available documents are too many, causing information overload also known as *infobesity* or *infoxication*. For these reasons, it is necessary to adopt automated tools to perform topics analysis favoring their understanding and the support to the decision making [1]. Digitalized sources have several characteristics such as heterogeneity and inconsistency in texts, which pose serious limits when trying to extract the meaning from sentences [2]–[4].

The real-time analyses support both companies in verifying the effectiveness of their marketing strategies and national or government agencies in evaluating political preferences and mood, or in supporting the earthquake and

disease information diffusion. These analyses can also be applied to other similar real-time decision-making environments, such as crisis management and emergency medical assistance [5]–[7], and not only to make decisions but also to simply describe the emotions of people, like suffering and hope, in the disaster-hit areas[8]. These analyses are allowed by the diffusion of data mining and machine learning techniques [9]–[11] due to the new devices, new software[12], [13], networking[14], [15] and new technologies [16]–[20] for new challenges such as to predict health states [21]–[23] or to building a virtual assistant health coach[24].

In textual format data, the automatic recognition process plays an important role. Information Extraction and Knowledge Discovery are key research areas focused on data analysis coming from different points of view connected to sources' heterogeneity[25]. For example, the information extraction in web data is complex for their dynamic nature [26], [27]. Here, there are changes for every new transaction and web usage, as well as for every field of applications. Therefore, one of the significant tasks for web applications is to collect users' opinions from different

sources to make this content available after the extraction process both for users and for social network applications, which can benefit from it for personalization and adaptation to user needs [28], [29]. Moreover, in this age of digital information overload, this content must be reliable for users and, hence, information sources must guarantee credibility [30]–[32] through the analysis of their context[33], semantic environment [34] and parameterized constraints[35]. A growing number of people are using social media for searching for information related to other sources. In this field, analysis of human behavior according to their textual contents is called opinion mining. Its purpose is to generate opinions from the data in the textual form [36], [37].

Opinion mining (also called sentiment analysis) [38] involves the creation of a system to collect and categorize opinions about a product or a service. Many organizations invest money and resources to search for opinions and make sentiment analysis. Sentiment analysis is also used to extract and classify opinions from the web, thus preventing a large amount of information on the web remains unused [39].

Sentiment analysis tries to capture the author's emotion and what he/she wants to communicate. Tools for sentiment analysis should be adapted to their use and to persons who use them. In the beginning, the sentiment analysis was used to analyze the opinions that people expressed on blogs. Turney [40] uses unsupervised classification methods based on reviews of cars, travels, and banks, thus classifying the considered topic as recommended or not recommended. In [41] the authors use stars as a reference scale.

Among social media, Twitter emerged as the most popular micro-blogging platform, where information proliferates rapidly and posted information and actions cause instantaneous responses from users [42]. Several researchers are discussing how to overcome the limitations of microblog-texts to recognize relevant tweets and extract, annotate and exploit domain-independent, semantic information[25], [43], [44]. Some other researchers are investigating how to exploit machine learning techniques like deep learning [9] in various cases, e.g., in the case of companies that want to know their customers better and opportunely manage the relationships with them, in the case of Arabic sentiment analysis on Twitter [45]. In the case of environments, such as in customer relationship management needed for the company to know their customer more closed [46]. In terms of Arabic sentiment analysis on Twitter [45] or recommending not only items of interest to the user, but also the conditions are enhancing user experiences with those items [47] are also investigated. Other researchers used naïve Bayes as a text classification algorithms to improve the accuracy of email and Twitter classification [48]. Twitter more than other social networks favors the sentiment analysis due to its “public” nature and offers some APIs<sup>1</sup> that allow displaying the tweets of users, to get the followers and following, to search for tweets by content, etc. The message shortness due to the limited number of characters on Twitter eases the analysis. Authors in [49] highlight the reasons to choose Twitter.

In this paper, we developed a software able to analyze a short text (i.e., the tweet in our case) and to identify the

sentiment expressed by the tweets posted on the Twitter social platform. According to the analysis we try to capture the orientation (e.g., judgment, thought, feeling or opinion) of the person who wrote the message. The software exploiting the language R is based on a Bayesian approach to classify the text. After a pre-processing (i.e., filtering of useless text and words), the developed software classifies the tweets expressing their sentiment (in positive, negative or neutral) and in an aggregated form. The software has been validated by identifying the minimum training set of data to be used to train the classifier. It is important to identify the minimum data set, which gives the best performance since the training phase is time-consuming. Then, the classifier has been used to classify real tweets on the work of U.S. President Donald Trump. The output accuracy (or error) of the software is evaluated and compared to the data provided by the Real Clear Politics website. It is worthy to note that this software allows accessing to the free opinions (by Twitter), expressed almost in real time to a large audience. Moreover, we investigate the possibility to analyze the expressed sentiment on a geographical basis, evaluating the position of the users who posted the tweets.

The paper is organized as follows. In section II, we describe the most popular methods to perform the sentiment analysis. In Section III, we provide the organization of the performed sentiment analysis. In Section IV, we describe the software developed to perform the analysis. In Section V, we apply the developed software to a real use case such as the popularity of the U.S. President. Finally, in Section VI conclusions are drawn.

## II. MATERIAL AND METHOD

### A. Related Works

Research on political communication has demonstrated that The role of the Internet as a source of political information produces the best sphere for public expression [50]. Due to its instantaneous response and its usage easiness, twitter more than other social media spread rapidly. These characteristics are ideal for promoting political viewpoints, particularly during contentious election campaigns. Political impact on social media has been examined through political discourses, and the political discourses on Twitter have been explored during the past several years by researchers. The emphasis on these studies was analyzing online networks of candidates and active users to forecast the results of the elections [51].

Different studies have investigated information behaviors of likeminded publics in the polarized political networks. Most of them focused on the strategic application of Twitter, highlighting the behaviors of the candidates while encouraging them to adopt Twitter with helpful tips for effective applications [52], [53]. Twitter can also be used to identify political preferences [54] and for day-by-day monitoring of electoral campaigns [55]–[57]. It is also possible to estimate the level of disaffection across people by counting negative tweets about politics in general, and this approach has shown that peaks in disaffection can correlate with important political news [58]. Other research studies explored how Twitter is applied within the electoral context to forecast electoral results [55], [59] by discovering

---

<sup>1</sup> <https://dev.twitter.com/>



#### D. Training Phase

In supervised machine learning techniques, a training phase is needed to create a knowledge base, where the system can learn useful information to make the classifications on another data subset [68]. In this paper, we adopted a Bayesian classifier. We report it in Fig. 2 the sequence diagram of the training phase, in which tweets are

filtered during the data filtering phase, and then two matrices called term-document matrix (tdm) are created: one for negative tweets and one for positive tweets. A document-term matrix or term-document matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents.

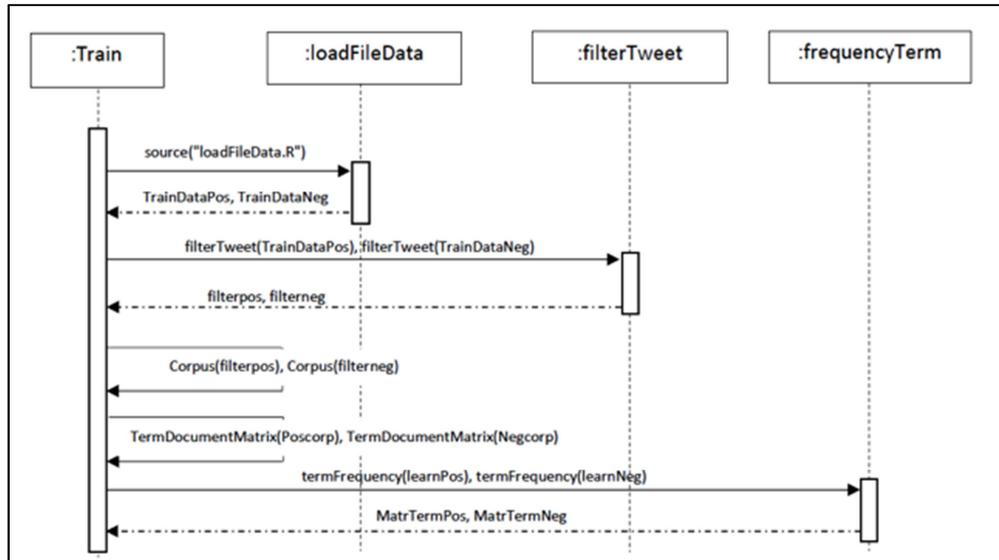


Fig. 2 Sequence Diagram of the Training phase

In a tdm, rows correspond to terms in the documents and columns correspond to documents in the collection. Then, the element  $E_{ij}$  of the tdm matrix is 1 if the term in row  $i$  is present in the document  $j$ , and 0 otherwise. Tdm matrix gives us information about the frequency of a term within a set of documents. Finally, the tdm matrix allows performing some data filtering operations, such as the deletion of numeric characters, punctuation or stop-words and length control of terms.

#### E. Classification Phase

A Bayesian classifier is used to classify the tweets. It divides the tweets into its relative tokens, and for each one, it calculates the probability that the tweet expresses positive or negative sentiment. For each term, the classifier provides the probability that the tweet has a positive polarity or negative polarity. Table 1 shows some terms with the relative frequency of appearance in positive and negative tweets.

TABLE I  
APPEARANCE FREQUENCY OF SOME TERMS

Term	Positive Tweet Frequency	Negative Tweet Frequency
Bad	1434	5228
Day	14045	10955
Good	16162	7570
Just	15599	15738
Love	10939	4038
Miss	306	6943

In Fig. 3 an example of the output generated by the classify function is reported. Percentages and absolute values of classification are shown in addition to the elapsed time required for classification.

	Positive	Negative	Neutral	TimeElapsed
Tot	592.000	370.000	35.0000	20.47
%	59.378	37.111	3.5105	0 m 20.47 s

Fig. 3 Example of output of the classify function

#### F. Validation Phase

The validation phase aims to evaluate the goodness of the classifier in terms of accuracy and error. The accuracy is defined as the ratio between the numbers of corrected classified tweets concerning the total analyzed tweets. Of course, the error is the complement to it, i.e.  $error = 1 - accuracy$ . To carry out this validation, the training set size is increased, and it is checked how the classifier performance varies.

After each training, a test phase follows where 20,000 tweets are classified (10,000 positives and 10,000 negatives). It starts with a training set of 2,000 tweets (1,000 positive and 1,000 negative) up to 160,000 tweets (80,000 positive and 80,000 negative). It is necessary that the tweets used for the training are different from those used for the test. In Fig. 4 the results of the accuracy are shown according to the size of the training set. We report the accuracy for the positive and negative class as well as its mean. Results show that the accuracy tends asymptotically to the value of 71% after 40,000 tweets.

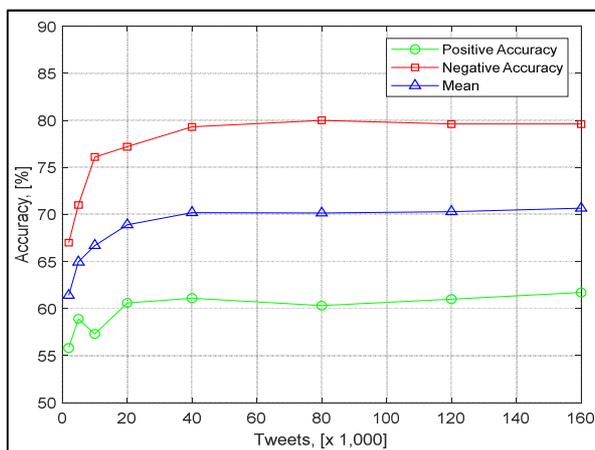


Fig. 4 Accuracy behavior vs. the training set increase

### G. Implementation

After evaluating the performance of the classifier on a pre-classified data set, we use the developed software on real data from the Twitter server. We used the “retweet” package to retrieve the tweets from the Twitter server. Moreover, it is necessary to create a new application on <https://apps.twitter.com>, defining its name, the “Consumer Key,” the “Consumer Secret Key” and some other information to use the `search_tweets` function to access the Twitter server. In this function, we set some parameters, such as the string to be searched among the tweets, the number of tweets to read and the language in which the tweets are written. Collected tweets are stored on the .csv file for the subsequent classification. The “tweets” object that is returned by the `search_tweets` function contains some useful information, such as the nickname of the user who wrote the tweet, the time the tweet was written, and the geographical coordinates of the device used to write the tweet.

### H. Visualizing Tweets with Their Geolocalization in R

Starting from the coordinates of the device used to write the tweet, we can visualize the location of the user on a map through the script. In Fig. 5 we reported an example of tweet positioning in the U.S. map. In figure 1, we also reported the analysis of the tweets by showing if the tweet is positive in green or negative in red. It is worthy to note that not all tweets contain geolocalization information. In Fig. 5, we analyzed 10,000 tweets, but we reported only 8%, containing device coordinates.

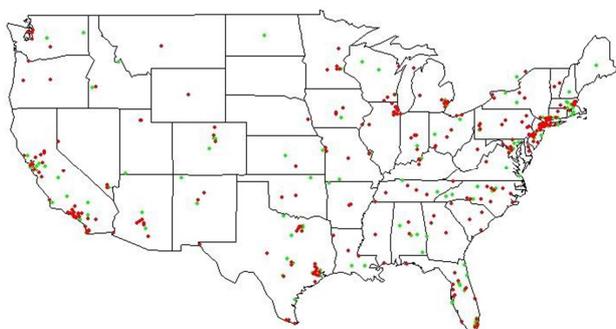


Fig. 5 Example of the positioning of positive (green) and negative (red) tweets in the U.S.

In addition to displaying the coordinates of the tweets on a map, we can go to retrieve the country of the coordinates of the tweet. The information on the country from where the tweet has been sent can be useful to evaluate the sentiment analysis about one topic (e.g., the popularity of the U.S. President in our case) based on the position of the user to be more effective in analyzing real-time people reactions. To this aim, we used the “revgeo” library enabling us to recover the country of the tweet’s user by his lat-lon.

## III. RESULTS AND DISCUSSION

In a society strongly driven by instinctive choices (i.e., not rational), it becomes increasingly important for politicians to know how they can gain more votes from potential people. Sentiment analysis tools can analyze in real time the reactions to certain debates and discussions, going to identify what are the issues that are most interesting for the population. In this section, we analyze a real use case using the developed Bayesian classifier. We investigate the general sentiment of some Twitter users about the work of the President of the United States Donald Trump, trying to carry out some indications on it.

### A. Reference Poll

To evaluate the performance of the classifier, it is necessary to have statistical data for comparison. Since we have not any survey, which refers to people’s feelings about the U.S. President work, we use statistical surveys on the work of President Trump in a given period. On the Internet, it is easy to find surveys of this type. Thus, we discovered tweets that contain the word “trump”, “Trump” or “TRUMP.” Then, we submitted them to the classifier. Finally, we compared the surveys published on the website [www.realclearpolitics.com](http://www.realclearpolitics.com). On Real Clear Politics there are surveys and statistics related to the U.S. political world. We referred to the period from 15<sup>th</sup> to 31<sup>st</sup> August 2018<sup>2</sup>, when 42.2% of the interviewees (on average) approved the work of Trump while 54.1% disapproved it.

### B. Classification

We adopted the script `miningTweet.R` to get a certain number of tweets containing the selected strings. We retrieved about 15,000 tweets per day for 5 days. Then, we fed the classifier with these tweets and the corresponding output (i.e. the behavior of positive, negative or neutral sentiment) in the different days is reported in Fig. 6. Comparing the output data of the classifier with those provided by *RealClear* Politics website shows that the percentages evaluated by the classifier in terms of approval and disapproval are similar to the results of the website. Fig. 6 shows that the output of the classifier provides 43.2% of positive sentiment, 47.1% of negative sentiment and about 9% of neutral sentiment. Further efforts can be done on tweets expressing neutral sentiment in order to have the same percentages.

<sup>2</sup> [https://www.realclearpolitics.com/epolls/other/president\\_trump\\_job\\_approval-6179.html](https://www.realclearpolitics.com/epolls/other/president_trump_job_approval-6179.html)

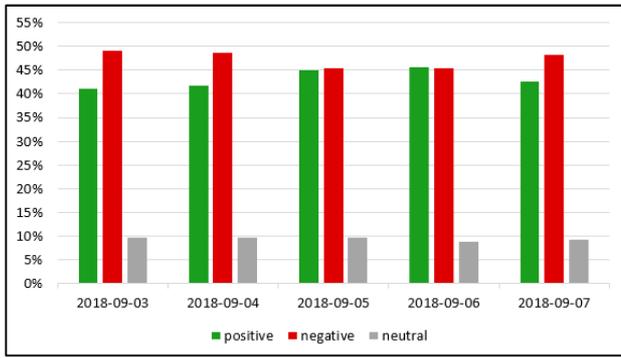


Fig. 6 collected tweets in five days with positive, negative or neutral sentiment

### C. Visualization and Geographic Localization

In this section, we report some examples showing the expressed sentiment on a map despite the low percentage of information regarding the location of tweets (lower than 10%). We used the ggmap library for these plots, through the get map function to retrieve the map and geom\_point function to draw the points on the retrieved map Fig. 7.



Fig. 7 Tweets showing the popularity of Trump work in U.S. map

This library allows the selection of a single Country but also of a particular city or an area to perform analyses on narrow geographical areas. In Fig. 8, we report an example of zooming the U.S. west coast, showing the geolocalized tweets and the related associated classification, as positive (in green) and negative (in red) ones.

Based on the information on the latitude and longitude of the tweet, it is possible to identify the country from where the tweet has been sent. We have aggregated the results by country and analyzed the difference between positive and negative geolocalized tweets on a country base.

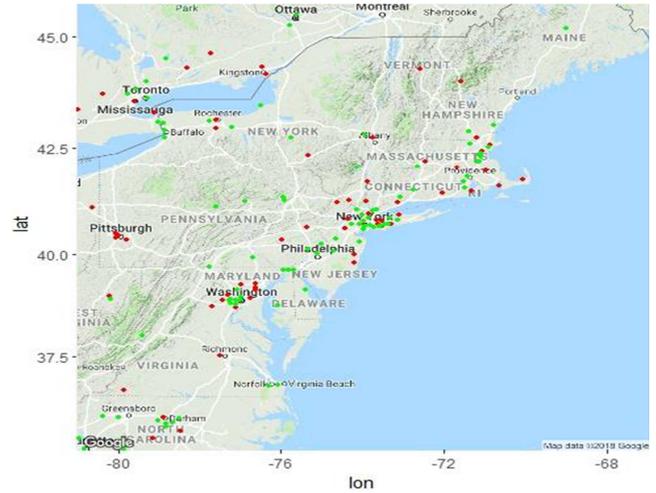


Fig. 8 Example of zoom in a restricted area showing classified tweets

In Fig. 9, we show the results of the sentiment distribution organized per country. Results are normalized to the maximum value given by country aggregate process. It is reported a negative (positive) sentiment if most of tweets of that state are negative (positive).

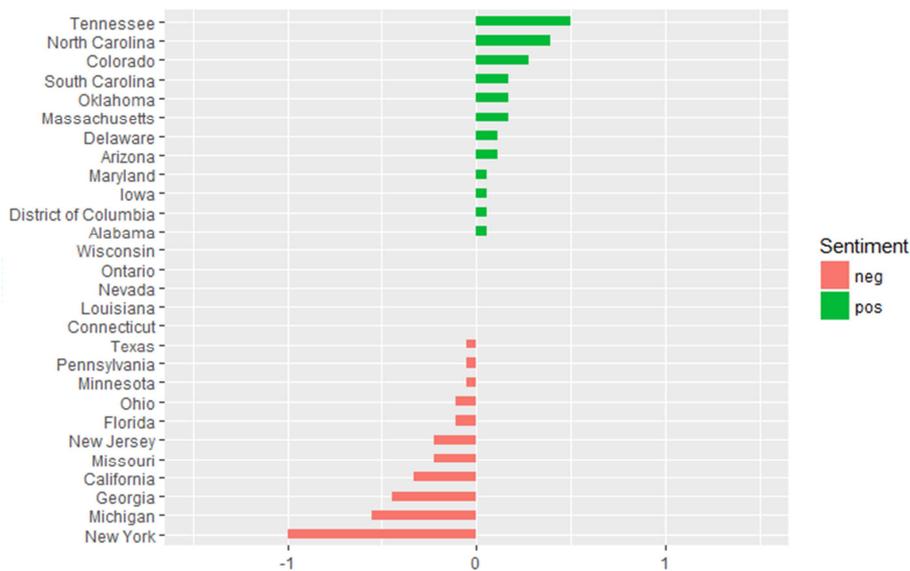


Fig. 9 Diverging Sentiment State Plot (Normalized Trump Sentiment)

#### IV. CONCLUSIONS

Social networks are a source of information with no rivals in terms of amount and variety of information that can be extracted from them. In particular, in this paper, we focus the attention on the possibility of identifying and collecting the free thoughts and opinions expressed by the people through social media. In this way, we can give as feedback to common people or to whom it may concern, e.g., politicians or technicians, the opportunity to read and interpret the average thought on a given topic. Here, we propose an approach to analyze, with the help of automated tools, comments, and opinions taken from social media in a real-time environment. The tool identifies sentiments in terms of positive, negative, neutral polarity expressed by the tweets posted on the U.S. President popularity. We perform the training phase with 40,000 tweets, and the classifier provides an average accuracy of about 70%. Then, the classifier has been used to classify real tweets on the work of U.S. President Donald Trump. The accuracy of the tool is similar to data provided by the Real Clear Politics website. The slight deviation (about 9% of neutral tweets evaluated by the classifier respect to about 4% of the Real Clear Politics website) is due to the reference period, which is not the same but successive of some days. Finally, we provided a visualization based on the geographical analysis of the sentiments of the users who posted the tweets.

#### REFERENCES

- [1] F. Nasution, N. E. Nazira Bazin, - Daliyusmanto, and A. Zulfikar, "Big Data's Tools for Internet Data Analytics: Modelling of System Dynamics," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 3, p. 745, 2017.
- [2] G. Fei, Z. Chen, A. Mukherjee, and B. Liu, "Discovering Correspondence of Sentiment Words and Aspects," 2018, pp. 233–245.
- [3] Z. Ansari, M. F. Azeem, A. V. Babu, and W. Ahmed, "A Fuzzy Approach for Feature Evaluation and Dimensionality Reduction to Improve the Quality of Web Usage Mining Results," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 2, no. 6, p. 477, 2016.
- [4] M. A. Taiye, S. S. Kamaruddin, and F. K. Ahmad, "Representing Semantics of Text by Acquiring its Canonical Form," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 3, p. 808, 2017.
- [5] F. Fallucchi, M. Tarquini, and E. W. De Luca, "Supporting Humanitarian Logistics with Intelligent Applications for Disaster Management," *INTELLI 2016*, p. 64, 2016.
- [6] D. Zhang, L. Zhou, and J. F. Nunamaker Jr, "A Knowledge Management Framework for the Support of Decision Making in Humanitarian Assistance/Disaster Relief," *Knowl. Inf. Syst.*, vol. 4, no. 3, pp. 370–385, 2002.
- [7] F. Fallucchi, M. Tarquini, and E. W. De Luca, *Knowledge management for the support of logistics during Humanitarian Assistance and Disaster Relief (HADR)*, vol. 265, 2016.
- [8] Y. Sato and I. Waragai, "The Function of Religious Language in the Media: A Comparative Analysis of the Japanese, German and American Newspaper Coverage about the 2011 Great East Japan Earthquake and Tsunami," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 2, p. 406, Apr. 2017.
- [9] L. Zhang, S. Wang, and B. Liu, "Deep Learning for Sentiment Analysis: A Survey," Jan. 2018.
- [10] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. S. Maida, "Deep Learning in Spiking Neural Networks," Apr. 2018.
- [11] E. W. De Luca and A. Nürnbergger, "Using clustering methods to improve ontology-based query term disambiguation," *Int. J. Intell. Syst.*, vol. 21, no. 7, pp. 693–709, Jul. 2006.
- [12] L. Bracciale, P. Loreti, A. Detti, R. Paolillo, and N. B. Melazzi, "Lightweight Named Object: an ICN-based Abstraction for IoT Device Programming and Management," *IEEE Internet Things J.*, pp. 1–1, 2019.
- [13] A. Detti *et al.*, "Application of Information-Centric Networking to NoSQL databases: The spatiotemporal use case," in *2017 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, 2017, pp. 1–6.
- [14] A. Detti, L. Bracciale, P. Loreti, G. Rossi, and N. Blefari Melazzi, "A cluster-based scalable router for information-centric networks," *Comput. Networks*, vol. 142, pp. 24–32, Sep. 2018.
- [15] L. Bracciale, P. Loreti, and G. Bianchi, "Human time-scale duty cycle for opportunistic WiFi-based mobile networks," in *2013 24th Tyrrhenian International Workshop on Digital Communications - Green ICT (TIWDC)*, 2013, pp. 1–6.
- [16] G. C. Cardarilli, A. Cristini, L. Di Nunzio, M. Re, M. Salerno, and G. Susi, "Spiking neural networks based on LIF with latency: Simulation and synchronization effects," in *2013 Asilomar Conference on Signals, Systems, and Computers*, 2013, pp. 1838–1842.
- [17] S. Acciarito, A. Cristini, L. Di Nunzio, G. M. Khanal, and G. Susi, "An a VLSI driving circuit for memristor-based STDP," in *2016 12th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 2016, pp. 1–4.
- [18] G. M. Khanal *et al.*, "Synaptic behavior in ZnO-rGO composites thin film memristor," *Electron. Lett.*, vol. 53, no. 5, pp. 296–298, Mar. 2017.
- [19] S. Acciarito *et al.*, "Hardware design of LIF with Latency neuron model with memristive STDP synapses," *Integration*, vol. 59, pp. 81–89, Sep. 2017.
- [20] A. Pieroni, N. Scarpato, L. Di Nunzio, F. Fallucchi, and M. Raso, "Smarter City: Smart energy grid based on Blockchain technology," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 1, 2018.
- [21] H. M. Rahman, N. Arbaiy, M. S. Che Lah, and N. Hassan, "Exploratory Study of Kohonen Network for Human Health State Classification," *JOIV Int. J. Informatics Vis.*, vol. 2, no. 3–2, p. 209, Jun. 2018.
- [22] J. Mir *et al.*, "Comparative Analysis for Heart Disease Prediction," *JOIV Int. J. Informatics Vis.*, vol. 1, no. 4–2, p. 227, 2018.
- [23] K. Madadipouya, "A Survey on Data Mining Algorithms and Techniques in Medicine," *JOIV Int. J. Informatics Vis.*, vol. 1, no. 3, p. 61, Jun. 2017.
- [24] I. Gupta *et al.*, "Towards Building a Virtual Assistant Health Coach," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 419–421.
- [25] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, "Challenges of Sentiment Analysis in Social Networks," in *Sentiment Analysis in Social Networks*, Elsevier, 2017, pp. 1–11.
- [26] F. Fallucchi, E. Alfonsi, A. Ligi, and M. Tarquini, *Ontology-driven public administration web hosting monitoring system*, vol. 8842, 2014.
- [27] F. Fallucchi, M. Petito, and E. W. De Luca, "Analysing and Visualising Open Data within the Data and Analytics Framework," Springer, Cham, 2019, pp. 135–146.
- [28] F. Cena, A. Dattolo, E. W. De Luca, P. Lops, T. Plumbaum, and J. Vassileva, "Semantic Adaptive Social Web," in *Proceedings of the 19th international conference on Advances in User Modeling*, Springer-Verlag, 2012, pp. 176–180.
- [29] T. Plumbaum, S. Wu, W. De Luca, Ernesto, and S. Albayrak, "User modeling for the social semantic web," *Proceedings of the Second International Conference on Semantic Personalized Information Management: Retrieval and Recommendation - Volume 781*. CEUR-WS.org, pp. 78–89, 2011.
- [30] M. Viviani and G. Pasi, "Credibility in social media: opinions, news, and health information-a survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 7, no. 5, p. e1209, Sep. 2017.
- [31] S. Y. Syn and S. U. Kim, "The impact of source credibility on young adults' Health information activities on Facebook: Preliminary findings," *Proc. Am. Soc. Inf. Sci. Technol.*, vol. 50, no. 1, pp. 1–4, Jan. 2013.
- [32] K. Ab Kadir, N. Sahari @ Ashaari, and J. Salim, "Credibility Dimensions for Islamic Information in Social Media," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 5, p. 1864, Oct. 2018.
- [33] K. Bade, E. W. De Luca, A. Nürnbergger, and S. Stober, "CARSA – An Architecture for the Development of Context Adaptive Retrieval Systems," Springer, Berlin, Heidelberg, 2006, pp. 91–101.
- [34] I. Korkontzelos, T. Zesch, T. U. Darmstadt, F. M. Zanzotto, and C. Biemann, "SemEval-2013 task 5: Evaluating phrasal semantics," *Proc. 7th Int. Work. Semant. Eval. (SemEval)*, vol. 2, no. SemEval, pp. 39–47, 2013.

- [35] M. Bianchi, M. Draoli, F. Fallucchi, and A. Ligi, "Service level agreement constraints into processes for document classification," in *ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems*, 2014, vol. 1.
- [36] B. Liu, *Sentiment analysis and opinion mining*. Morgan & Claypool, 2012.
- [37] S. N. Khan, N. Mohd Nawi, M. Imrona, A. Shahzad, A. Ullah, and A.- Rahman, "Opinion Mining Summarization and Automation Process: A Survey," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 5, p. 1836, Oct. 2018.
- [38] B. Liu, *Sentiment analysis: mining opinions, sentiments, and emotions*.
- [39] S. Saad and B. Saberi, "Sentiment Analysis or Opinion Mining: A Review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, p. 1660, 2017.
- [40] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. 40th Annu. Meet. Assoc. Comput. Linguist.* No. July, pp. 417–424, 2002.
- [41] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. Vol. 2, no. No 1-2, pp. 1–135, 2008.
- [42] M. Chong, "Analyzing political information network of the U.S. Partisan public on twitter," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10766 LNCS, pp. 453–463, 2018.
- [43] S. Narr, E. W. De Luca, and S. Albayrak, "Extracting semantic annotations from twitter," in *Proceedings of the fourth workshop on Exploiting semantic annotations in information retrieval - ESAR '11*, 2011, p. 15.
- [44] A. Said, S. Berkovsky, and E. W. De Luca, "Introduction to special section on CAMRa2010," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 1, pp. 1–9, Jan. 2013.
- [45] M. M. Altawaier and S. Tiun, "Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, p. 1067, 2016.
- [46] H. Amnur, "Customer Relationship Management and Machine Learning Technology for Identifying the Customer," *JOIV Int. J. Informatics Vis.*, vol. 1, no. 4, p. 12, 2018.
- [47] S. Aciar, D. Zhang, S. Simoff, and J. Debenham, "Informed Recommender: Basing Recommendations on Consumer Product Reviews," *IEEE Intell. Syst.*, vol. 22, no. 3, pp. 39–47, May 2007.
- [48] A. Erianda and I. Rahmayuni, "Improvement of Email And Twitter Classification Accuracy Based On Preprocessing Bayes Naive Classifier Optimization In Integrated Digital Assistant," *JOIV Int. J. Informatics Vis.*, vol. 1, no. 2, p. 53, 2018.
- [49] R. Heimann and N. Danneman, *Social Media Mining with R*. Birmingham: Packt Publishing, 2014.
- [50] D. V. Shah, J. Cho, W. P. Eveland, and N. Kwak, *Information and expression in a digital age: Modeling internet effects on civic participation*, vol. 32, no. 5. 2005.
- [51] A. Makazhanov and D. Rafiei, "Predicting political preference of Twitter users," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, 2013, pp. 298–305.
- [52] R. H. Lasseter, "MicroGrids," *2002 IEEE Power Eng. Soc. Winter Meet. Conf. Proc. (Cat. No.02CH37309)*, vol. 1, pp. 305–308, 2002.
- [53] C. B. Williams and G. J. Gulati, "Communicating with Constituents in 140 Characters or Less : by," *Polit. Sci.*, 2010.
- [54] J. Golbeck, C. Park, and D. L. Hansen, "Computing Political Preference among Twitter Followers," *Hum. Factors*, pp. 1105–1108, 2011.
- [55] A. Ceron, L. Curini, and S. M. Iacus, "Using Social Media To Forecast Electoral Results: a Review of the State of the Art," *Ital. J. Appl. Stat.*, vol. 25, no. 3, 2014.
- [56] M. J. Jensen and N. Anstead, "Psephological investigations: Tweets, votes, and unknown unknowns in the republican nomination process," *Policy & Internet*, vol. 5, no. 2, pp. 161–182, Jun. 2013.
- [57] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle," *Proc. 50th Annu. Meet. Assoc. Comput. Linguist.*, no. July, pp. 115–120, 2012.
- [58] C. Monti, A. Arvidsson, A. Rozza, M. Zignani, E. Colleoni, and G. Zappella, "Modelling political disaffection from Twitter data," pp. 1–9, 2013.
- [59] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas, "More tweets, more votes: Social media as a quantitative indicator of political behavior," *PLoS One*, vol. 8, no. 11, pp. 1–11, 2013.
- [60] T. Graham, M. Broersma, K. Hazelhoff, and G. van 't Haar, *Between Broadcasting Political Messages and Interacting With Voters*, vol. 16, no. 5. 2013.
- [61] A. Bruns and T. Highfield, "POLITICAL NETWORKS ON TWITTER: Tweeting the Queensland state election," *Inf. Commun. Soc.*, vol. 16, no. 5, pp. 667–691, 2013.
- [62] "Push Attack: Binding Virtual and Real Identities Using Mobile Push Notifications," *Futur. Internet*, vol. 10, no. 2, p. 13, Jan. 2018.
- [63] D. Gayo-Avello, "No, you cannot predict elections with twitter," *IEEE Internet Comput.*, vol. 16, no. 6, pp. 91–94, 2012.
- [64] M. A. Razzaq, A. M. Qamar, and H. S. M. Bilal, "Prediction and analysis of Pakistan election 2013 based on sentiment analysis," *ASONAM 2014 - Proc. 2014 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, no. Asonam, pp. 700–703, 2014.
- [65] M. Broersma and T. Graham, "Tweets as news source during the 2010 British and Dutch elections," *Journal. Pract.*, vol. 6 (3), no. June 2010, pp. 403–419, 2012.
- [66] D. Pace, H. Thayer, and H. Thayer, "21st Century Propaganda: The Age of Twitter," 2018.
- [67] N. Idris, S. Z. Mohd Hashim, R. Samsudin, and N. B. Hj Ahmad, "Intelligent Learning Model Based On Significant Weight Of Domain Knowledge Concept For Adaptive E-Learning," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 4–2, p. 1486, Sep. 2017.
- [68] T. M. Mitchell, *Thomas Mitchell-Machine learning-McGraw Hill Higher Education (1997)*.