Spatio-Temporal Trend Analysis of the Brazilian Elections based on Twitter Data

Bruno J. G. Praciano[1], João Paulo C. L. da Costa[1], João Paulo A. Maranhão[1], Fabio L. L. de Mendonça[1], Rafael T. de Sousa Junior[1], and Juliano B. Prettz[1]

[1]*Department of Electrical Engineering, University of Brasília (UnB), Brasília, Brazil*

***Abstract*—Text classification techniques and sentiment analysis can be applied to understand and predict the behavior of users by exploiting the massive amount of data available on social networks. In this context, trend analysis tools based on supervised machine learning are crucial. In this work, a framework for spatio-temporal trend analysis of Brazilian presidential elections based on Twitter data is proposed. Experimental results show that the proposed framework presents good effectiveness in predicting election results as well as providing tweet author's geolocation and tweet timestamp, with an accuracy close to 90% when the Support Vector Machine (SVM) algortihm is applied for sentiment classification.**

***Index Terms*—Big Data, Trend Analysis, Supervised Machine Learning, Support Vector Machine**

I. Introduction

Social networks generate a large amount of data which is openly and easily available on the Internet [1]. One example of online social network is the Twitter, whose users can express opinions with a limited amount of characters in real time. This information can be exploited by machine learning algorithms in order to support the trend analysis [2]. To perform such trend analysis, text classification techniques and sentiment analysis are crucial [3].

Dictionaries are fundamental for the text classification and sentiment analysis. Therefore, one challenge is to perform sentiment analysis in other languages than English. In this sense, in order to obtain an accurate analysis, it is necessary to use a native dictionary to perform the classification of texts in an automated fashion. The text classification is performed by using supervised machine learning algorithms [4].

In this paper we propose a framework for trend analysis of Brazilian presidential election based on Twitter data sentiment analysis. First, an application for crawling and data extracting from Twitter database is developed. Next, the raw information is pre-processed in order to standardize tweet texts and clean the data. Then, sentiment analysis is performed on the data obtained from the previous step through the applying of data mining algorithms. Finally, we perform geolocation and timestamp extracting where information about the geographic coordinates of the tweet author as well as date and time are obtained. The proposed framework is validated through the comparison with the 2014 Brazil presidential election results extracted from the Superior Electoral Court database. According to our results, the proposed framework shows good effectiveness in predicting election results and providing tweet author's geolocation and tweet timestamp.

The remainder of this paper is organized as follows. Section II presents the related works. Section III describes the proposed framework for trend analysis of Brazil presidential election based on Twitter data sentiment analysis. In Section IV the simulation results are presented and discussed. Finally, Section V concludes the paper.

II. Related Works

In [1] the authors applied text-mining techniques to Twitter data related to the 2012 Korean presidential election. Three primary techniques were used: topic modeling to track changes in topical trends, mention-direction-based user network analysis, and term co-occurrence retrieval for further content analysis. The results show that Twitter is effective to detect and trace changes in social issues.

In [5] it is proposed a system which surveys the French presidential election trends from Twitter discussions through the analysis of polarity and intensity of opinion. Another objective of the authors consisted of searching and locating the content corresponding to the 2012 French presidential election in posted tweets. The results show the convergence of the obtained results with the official polling statistics, associated with the change in popularity of candidates after their election speeches during the campaign.

In [6], the authors collected a dataset related to 2014 Colombian presidential election tweets and a supervised learning technique was implemented on a labeled collection of users in order to distinguish spammer accounts from non-spammer ones. They developed and applied a sentiment analysis system aiming to investigate the potential of social media for voting intention inference. According to the experimental results, inference methods based on Twitter data are not consistent, despite obtaining the lowest mean absolute error and correctly ranking the highest-polling candidates in the first round election with the proposed inference method.

In [7] it was used a lexicon and Naïve Bayes machine learning algorithm to calculate the sentiment of political tweets collected 100 days before the 2016 US presidential election. The authors used manually and automatically labeled tweets based on hashtag content/topic. The results suggested that Twitter is becoming a more reliable platform in comparison to previous works.

2375-9259/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDMW.2018.00192

In [3] the SVM algorithm combined with the Sentilex Dictionary classifies sentiments of tweets for political trend analysis. The work used the dataset of the 2010 Brazilian presidential elections and the achieved accuracy was 81,37%.

To the best of our knowledge, our work is the first one to propose a framework for spatio-temporal sentiment analysis considering a Brazilian Portuguese dictionary. We evaluate the performance of the sentiment analysis comparing the following machine learning algorithms: SVM, Naive Bayes, decision trees and logistic regression.

III. Proposed Framework for Spatio-Temporal Trend Analysis Based on Twitter Data

In this section, we detail our proposed framework for trend analysis of the Brazilian elections based on Twitter data sentiment analysis. As shown in Figure 1, the framework is divided into functional blocks, which are described in Subsections III-A to III-D.

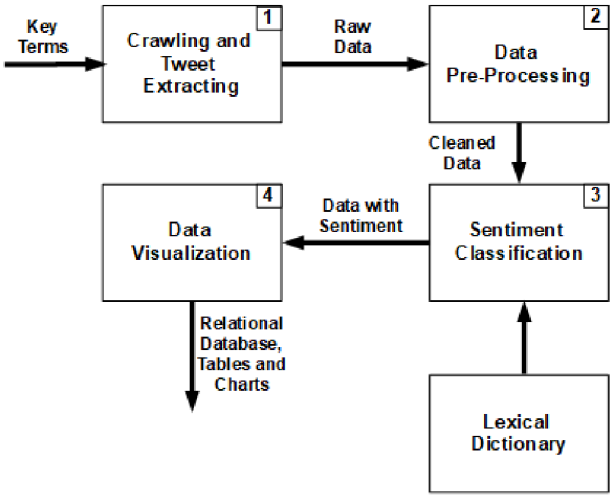

Fig. 1. Block diagram of the proposed framework for spatio-temporal trend analysis of the Brazilian elections based on Twitter data

A. Crawling and Tweet Extracting

The first block of Figure 1 corresponds to the crawling and tweet extracting. In this block, tweets are extracted from the Twitter database. Since the new version of the Twitter API does not allow to extract tweets older than a week, it was necessary to develop a Web Scrapy application. Besides the tweet text, addition information is collected from the extracted tweets including author's name, date, count of retweets, count of favourites, ID and link of publication, localization, mentions and hashtags.

B. Data Pre-Processing

The second block of Figure 1 corresponds to the data pre-processing. There is no defined writing pattern to be used on social networks. Consequently it was necessary to perform data cleaning in order to standardize the sentences. In addition, all emoticons were removed since we intend to analyze the lexicon of the Portuguese language. The internal flow of the data pre-processing block 2 is described in Figure 2.

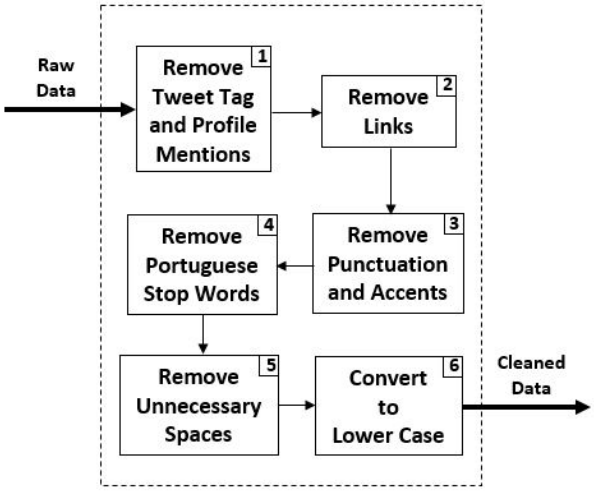

Fig. 2. Internal flow of the data pre-processing block

We defined six steps to perform the data cleaningWe defined six steps to perform the data cleaning, as described in 2. The first block is responsible for removing tweet tags and profile mentions. Next, the second and third blocks remove web site links, as well as accents and punctuations, respectively. Then all Portuguese stop words and unnecessary spaces are removed in the fourth and fifth blocks, respectively. Finally, the sixth block converts the tweet text to lower case letters.

Stop words play a very small role in sentiment analysis and consequently must be removed [8]. They do not contribute for the sentiment analysis process and only slow the process down. Moreover, the data must be submitted to a stemming process in which the words are reduced to their stem, i.e., their root form [3].

C. Sentiment Classification

The third block of Figure 1 corresponds to the sentiment classification of the tweets after their pre-processing. Sentiment analysis allows us to classify sentences as positive, negative or neutral [5].

The text classifier must be trained with a training dataset $(\mathbf{F}_i, c_i)$, where $\mathbf{F}_i$ is the feature vector and $c_i = \{+1, -1\}$ is the class index corresponding to the i-th tweet. The training dataset is obtained by applying the traditional text classification approach known as Bag-of-Words (BoW) [9]. This model represents a text as an unordered collection of its words, disregarding grammar and word order. For text classification, a weight is assigned to each word according to its frequency in the document [10].

The classifier is trained by finding the discriminant function g($\cdot$) such that if g($\mathbf{F}_i \geq t_2$, then c_i = positive, if $t_1 \leq$ g($\mathbf{F}_i$) $\leq t_2$, then c_i = neutral or if g($\mathbf{F}_i$) $\leq t_1$, then c_i = negative, where t_1 and t_2 are thresholds.

In this work we use the libraries TextBlob [11] and OpLexicon [12] combined with Sentilex [13] for processing textual data. The following state-of-the-art data mining algorithms are applied for sentiment classification: Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR) and Decision Trees (DT). Next the application of these techniques in our framework is described.

1) Support Vector Machine: According to [14], the SVM algorithm consists of computing, in the training phase, the maximum margin separation hyperplane in the feature space which can best separate the different classes. The hyperplane is completely determined by the unitary vector $\mathbf{a}$ normal to the hyperplane and the bias b corresponding to the perpendicular distance from hyperplane to the origin. Therefore, the training phase consists in determining $\mathbf{a}$ and b such that $g(\mathbf{F}_i) = \mathbf{a}^T\mathbf{F}_i + b$, where $\mathbf{a}^T\mathbf{F}_i + b = 0$ holds for points $\mathbf{F}_i$ that belongs to this maximum margin hyperplane.

The first step of SVM consists of the training phase where the unitary vector $\mathbf{a}$ and the bias b are obtained and consequently the maximum margin separation hyperplane is computed. After the training phase, the SVM classifier is obtained and finally the text data can be classified.

2) Naïve Bayes: The Naïve Bayes approach is the simplest and most commonly used classifier [15]. The first step of the Naive Bayes consists of the training phase, where the prior probabilities of each class $P(c_i)$ as well as the likelihood of different features f_i for each class $P(f_i|c_i)$ are learned. Next, assuming that all features f_i are independent, the NB algorithm computes the probability $P(c_i|\mathbf{F}_i)$ that a given feature set $\mathbf{F}_i$ belongs to a particular class label. Finally, c_i corresponds to the greatest probability value obtained in the previous step, i.e., $\underset{c_i}{\operatorname{argmax}}(P(c_i|\mathbf{F}_i))$.

3) Logistic Regression: The logistic regression approach is a generalized linear model that extends the linear regression model by linking the range of real numbers to the 0-1 range [16].

Given the feature set $\mathbf{F}_i$, the logistic regression algorithm computes the probability that the response variable takes on value 1 from the training dataset. The first step of the logistic regression consists of the training phase, where the weight coefficients β_i of the logistic response function are obtained. Next, the probability $\pi(x_i)$ that the response variable takes on value 1 is computed. Finally, the final classification is performed. For more detailed information, the reader is referred to [16].

4) Decision Tree: The decision tree provides a hierarchical decomposition of the data space in which a condition on the attribute value is used to divide the data [17]. The division of the data space is performed recursively until the leaf nodes contain certain minimum numbers of records which are used for the purpose of classification [15].

TABLE I
TWEET POLARITY USING TEXTBLOB AND OPLEXICON/SENTILEX LIBRARIES

	Positive	Neutral	Negative
TextBlob	42.67%	24.01%	33.27%
OpLexicon/Sentilex	25.12%	26.51%	48.35%

TABLE II
SENTIMENTAL ANALYSIS OF A SAMPLE WITH THREE TWEETS

Tweet	TextBlob	OpLexicon/Sentilex
1	Negative	Positive
2	Neutral	Neutral
3	Negative	Negative

After the root node is created, the algorithm selects the best way to split the records based on the degree of impurity of the child nodes. The tree-growing process is finished after a stopping condition is satisfied and then a class label is assigned to the leaf node. Finally, the text can be applied on the DT classifier in order to obtain the classification c_i.

D. Data Visualization

The fourth block of Figure 1 corresponds to the data visualization. Note that the tweet author can make their geographic coordinates available in their personal profile. Therefore, it is possible to determine the geographic location of the users and consequently intensify the marketing actions to be taken in that region.

IV. RESULTS

This section presents the validation of the proposed framework. The framework trends are compared to the 2014 presidential election results extracted from the Brazilian Superior Electoral court database. We observed only the second round results of the presidential elections and, therefore, the candidates under analysis were Dilma Rouseff and Aécio Neves, who obtained, in the first round, the highest amount of votes.

Our dataset has 158,279 tweets which were collected from the Twitter database considering the period between October 12, 2014 and October 28, 2014 and the elections results were available at October 26, 2014.

Table I shows the polarity percentages for all extracted tweets after their pre-processing considering lexicon libraries [11] [12] [18]. A large difference can be observed between the results provided by both libraries, especially considering the positive and negative classifications.

Table II shows an example of the sentiment analysis of a random sample with three tweets considering the previous libraries. The sentiment classification presents the same results for tweets 2 and 3, while tweet 1 was respectively classified as negative and positive by using TextBlob and OpLexicon/Sentilex.

First, we tried to use the TextBlob library as the main source for classification. However, since TextBlob uses a lexicon dictionary of English words [19], it is necessary to translate to English all tweets written in other languages and then check

the phrase polarity. In this process, words may lose their meaning because the translation process adds errors.

As shown in Table III, by using OpLexicon library, the best results consisted of 30,322 words (23,433 adjectives and 6,889 verbs) and were based on Brazilian Portuguese. It was classified by its morphological category labelled with polarities positive, negative and neutral.

The Sentilex library is a lexicon of sentiments for Portuguese constituted by 6,321 adjectival lemmas (by convention, in the singular masculine form) and 25,406 flexed forms. In this library, the polarity can be positive, negative or neutral.

A. Performance evaluation of the trend analysis

The evaluation of sentiment classification is performed by using the following metrics: accuracy, precision, recall and F1 score [20]. These indexes are given by:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{4}$$

where TP, FN, FP and TN are, respectively, the number of true positive instances, false negative instances, false positive instances and true negative instances.

Table III shows the performance of SVM, NB, LR and DT classification algorithms by using two different libraries. SVM algorithm presented the best results and the lowest computational cost. Text data are ideally suited for SVM classification due to the sparse nature of text, in which few features are irrelevant, but they tend to be correlated with one another and generally organized into linearly separable categories [15].

B. N-Fold Cross Validation

In this subsection, we show the N-fold cross-validation of the framework results. The dataset is partitioned into N mutually exclusive subsets. One subset is used as validation data for testing the model, and the remaining $N-1$ subsets are used as training data. Since the dataset has more than 100,000 tweets, the N-fold cross-validation would present a high computational cost whether a great value of N was chosen. Consequently, we used $N = 5$. Table IV illustrates the results for 5-fold cross validation considering TextBlob and OpLexicon/Sentilex libraries. Once again, SVM algorithm presented the best results for all evaluated metrics.

TABLE III
ALGORITHM PERFORMANCE AND RESPECTIVE DICTIONARIES

Metrics	Algorithm	TextBlob	OpLexicon/Sentilex
Accuracy	NB	0.82	0.93
	SVM	0.94	0.98
	LR	0.70	0.65
	DT	0.64	0.85
Precision	NB	0.83	0.92
	SVM	0.94	0.98
	LR	0.79	0.83
	DT	0.69	0.89
Recall	NB	0.79	0.92
	SVM	0.93	0.97
	LR	0.60	0.58
	DT	0.56	0.81
F1 score	NB	0.80	0.92
	SVM	0.94	0.98
	LR	0.55	0.60
	DT	0.55	0.84

TABLE IV
5-FOLD CROSS VALIDATION

Metrics	Algorithm	TextBlob	OpLexicon/Sentilex
Accuracy	NB	0.81	0.92
	SVM	0.93	0.98
	LR	0.86	0.66
	DT	0.64	0.84
Precision	NB	0.82	0.92
	SVM	0.92	0.98
	LR	0.78	0.83
	DT	0.70	0.88
Recall	NB	0.86	0.94
	SVM	0.93	0.98
	LR	0.90	0.80
	DT	0.54	0.81
F1 score	NB	0.79	0.91
	SVM	0.92	0.98
	LR	0.54	0.58
	DT	0.56	0.84

C. Error Evaluation of the Sentiment Analysis via SVM

Since SVM presented the best results among all data mining algorithms under analysis, it was applied for the error evaluation of the sentiment analysis. The training dataset consists of 3000 tweets which were divided into three groups of 1000 tweets each, where each group presents a given polarity (positive, neutral or negative). As described in the confusion matrix of Figure 3, our framework correctly classified the positive, negative and neutral tweets with rates of 99.50%, 86.90% and 70.60%, respectively. In this case we used the key terms "Aécio" and "Dilma", which refer to the candidate names Aécio Neves and Dilma Rousseff.

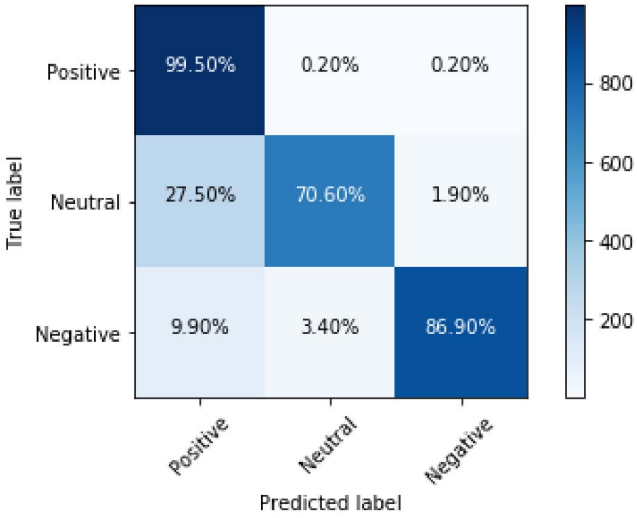

Fig. 3. Confusion matrix for manual classification with SVM considering the key terms "Dilma" and "Aécio" and 3000 tweets in total

D. Spatio-Temporal Trend Analysis

In the traditional temporal trend analysis, the sentiment analysis is performed on the total data from different users in different locations. On the other hand, in a spatial analysis, the data of each location is separately estimated, which provides a higher accuracy in our trend analysis.

Figure 4 shows the spatio-temporal trend analysis based on classified tweets during the vote counting of Dilma Rousseff and Aécio Neves for all 26 states and the Federal District of Brazil, considering the period between October 12, 2014 and October 28, 2014. The candidate with a higher percentage of positive tweets is assumed to be selected by the location

Figures 5 and 6 illustrates respectively the intensity maps for the 2014 presidential election results provided by the Superior Electoral Court and the predictions provided by the proposed framework. According to those Figures, there are seven states with wrong trends, namely Acre, Roraima, Amazonas, Amapá, Tocantins, Piauí and Mato Grosso do Sul. Approximately 6.5 million people voted for the 2014 presidential elections in those states, while about 99.0 million voted at the remaining

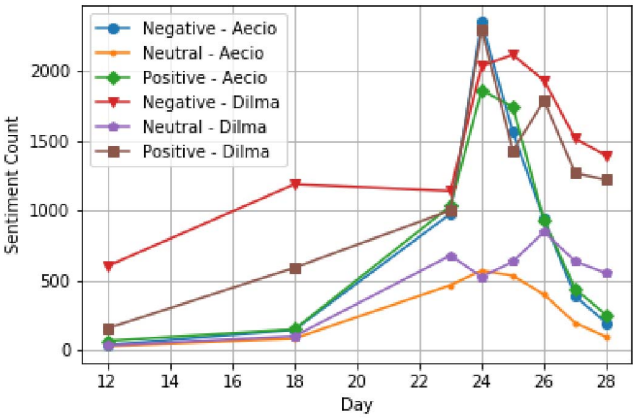

Fig. 4. Temporal analysis based on classified tweets during the vote counting of Dilma Roussef and Aécio Neves

states and the Federal District. Therefore, the percentage of error considering a weighted spatial distribution is about 6.5

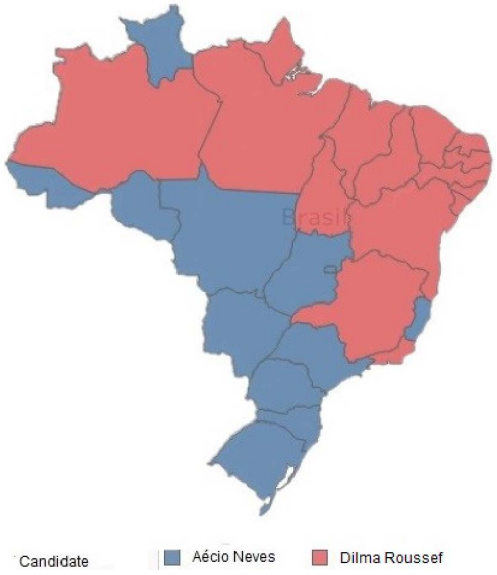

Fig. 5. Election results provided by the Superior Electoral Court [21]

V. Conclusion and Future Works

In this work we propose a framework for spatio-temporal trend analysis of Brazilian presidential elections based on Twitter data. According to our results, an accuracy close to 90% is obtained when the SVM algorithm is applied for sentiment classification.

For future works we expect to include a greater amount of words in our dictionary in order to increase the framework reliability. We also intend to analyze more sentiments, use deep learning to extract additional information as well as use pre-trained word vectors aiming to obtain a higher performance.

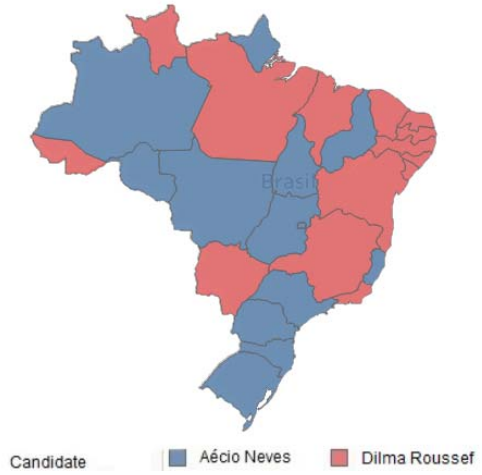

Fig. 6. Election prediction provided by the proposed framework

ACKNOWLEDGMENT

The authors would like to thank the Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF) for their funding under the calls 03/2015 and 01/2017, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the PNPD, PVE and PDE scholarships, and CAPES/PROBAL Proc. 88887.144009/2017-00. João P. C. L. da Costa is partially supported by the National Council for Scientific and Technological Development – CNPq.

This research work has the support of the Brazilian research and innovation Agencies CAPES - Coordination for the Improvement of Higher Education Personnel (Grant 23038.007604/2014-69 FORTE - Tempestive Forensics Project), CNPq – National Council for Scientific and Technological Development (Grant 465741/2014-2 Science and Technology National Institute – INCT on Cybersecurity) and FAPDF - Research Support Foundation of the Federal District (Grants 0193.001366/2016 UIoT - Universal Internet of Things and 0193.001365/2016 - Secure Software Defined Data Center - SSDDC), as well as the Ministry of Planning, Development and Management (Grants 005/2016 DIPLA - Planning and Management Directorate, 011/2016 SEST - Secretariat of State-owned Federal Companies, and 083/2016 ENAP - Brazilian Public Administration School), the Institutional Security Office of the Presidency of the Republic of Brazil (Grant 002/2017), and the DPGU – Brazilian Union Public Defender (Grant 066/2016).

REFERENCES

[1] M. Song, M. C. Kim, and Y. K. Jeong, "Analyzing the political landscape of 2012 Korean presidential election in Twitter," *IEEE Intelligent Systems*, vol. 29, no. 2, pp. 18–26, 2014.

[2] E. Rodrigues, R. Assunção, G. L. Pappa, D. Renno, and W. M. Jr, "Exploring multiple evidence to infer users' location in Twitter," *Neurocomputing*, vol. 171, pp. 30–38, 2016.

[3] D. Tumitan and K. Becker, "Sentiment-based features for predicting election polls: a case study on the Brazilian scenario," in *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on*, vol. 2. IEEE, 2014, pp. 126–133.

[4] L. A. de Freitas and R. Vieira, "Exploring resources for sentiment analysis in Portuguese language," in *Intelligent Systems (BRACIS), 2015 Brazilian Conference on*. IEEE, 2015, pp. 152–156.

[5] K. Wegrzyn-Wolska and L. Bougueroua, "Tweets mining for French presidential election," in *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, Nov 2012, pp. 138–143.

[6] J. A. Cerón-Guzmán and E. León-Guzmán, "A sentiment analysis system of Spanish tweets and its application in Colombia 2014 presidential election," in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, Oct 2016, pp. 250–257.

[7] B. Joyce and J. Deng, "Sentiment analysis of tweets for the 2016 US presidential election," in *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*, Nov 2017, pp. 1–4.

[8] P. Sharma and T. S. Moh, "Prediction of Indian election using sentiment analysis on hindi Twitter," in *Big Data (Big Data), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1966–1971.

[9] S. T. Indra, L. Wikarsa, and R. Turang, "Using logistic regression method to classify tweets into the selected topics," in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct 2016, pp. 385–390.

[10] S. G. K and S. Joseph, "Text classification by augmenting bag of words (bow) representation with co-occurrence feature," vol. 16, pp. 34–38, 01 2014.

[11] S. Loria. TextBlob: Simplified text processing. [Online]. Available: http://textblob.readthedocs.io/en/dev/index.html

[12] M. Souza and R. Vieira, "Sentiment analysis on twitter data for portuguese language," in *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*, ser. PROPOR'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 241–247. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-28885-2_28

[13] B. Neuenschwander, A. C. Pereira, W. Meira, Jr., and D. Barbosa, "Sentiment analysis for streams of web data: A case study of brazilian financial markets," in *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*, ser. WebMedia '14. New York, NY, USA: ACM, 2014, pp. 167–170. [Online]. Available: http://doi.acm.org/10.1145/2664551.2664579

[14] P. M. G. I. Reis, J. P. C. L. da Costa, R. K. Miranda, and G. D. Galdo, "Esprit-hilbert-based audio tampering detection with svm classifier for forensic analysis via electrical network frequency," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 4, pp. 853–864, April 2017.

[15] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093 – 1113, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2090447914000550

[16] U. Ananthakumar and R. Sarkar, "Application of logistic regression in assessing stock performances," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, Nov 2017, pp. 1242–1247.

[17] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986. [Online]. Available: http://dx.doi.org/10.1023/A:1022643204877

[18] W. M. J. B. Neuenschwander, A. Pereira and D. Barbosa, "Sentiment analysis for streams of web data: A case study of Brazilian financial markets," in *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*. ACM, 2014, pp. 167–170.

[19] G. A. Miller, "Wordnet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[20] Z. Hailong, G. Wenyan, and J. Bo, "Machine learning and lexicon based methods for sentiment classification: A survey," in *2014 11th Web Information System and Application Conference*, Sept 2014, pp. 262–265.

[21] Tribunal Superior Eleitoral. Estatísticas Eleitorais: eleições anteriores. [Online]. Available: http://www.tse.jus.br/eleitor-e-eleicoes/estatisticas/eleicoes/eleicoes-anteriores/estatisticas-eleitorais-anos-anteriores