



Forecasting the Chilean Electoral Year: Using Twitter to Predict the Presidential Elections of 2017

Sebastián Rodríguez¹(✉), Héctor Allende-Cid¹, Wenceslao Palma¹,
Rodrigo Alfaro¹, Cristian Gonzalez², Claudio Elortegui³, and Pedro Santander³

¹ Escuela de Ingeniería Informática,
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
{sebastian.rodriguez.o,hector.allende,wenceslao.palma,
rodrigo.alfaro}@pucv.cl

² Instituto de Ciencias de Lenguaje y Literatura,
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
cristian.gonzalez@pucv.cl

³ Escuela de Periodismo,
Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
{claudio.elortegui,pedro.santander}@pucv.cl

Abstract. Failures of traditional survey methods for measuring political climate and forecasting high impact events such as elections, offers opportunities to seek alternative methods. The analysis of social networks with computational linguistic methods have been proved to be useful as an alternative, but several studies related to these areas were conducted after the event (post hoc). Since 2017 was the election year for the 2018–2022 period for Chile and, moreover, there were three instances of elections in this year. This condition makes a good environment to conduct a case study for forecasting these elections with the use of social media as the main source of Data. This paper describes the implementation of multiple algorithms of supervised machine learning to do political sentiment analysis to predict the outcome of each election with Twitter data. These algorithms are Decision Trees, AdaBoost, Random Forest, Linear Support Vector Machines and ensemble voting classifiers. Manual annotations of a training set are conducted by experts to label pragmatic sentiment over the tweets mentioning an account or the name of a candidate to train the algorithms. Then a predictive set is collected days before the election and an automatic classification is performed. Finally the distribution of votes for each candidate is obtained from this classified set on the positive sentiment of the tweets. Ultimately, an accurate prediction was achieved using an ensemble voting classifier with a Mean Absolute Error of 0.51% for the second round.

Keywords: Election forecasting · Sentiment analysis
Machine learning · Ex Ante forecasting

1 Introduction

2016 was a year in which two classic institutions undoubtedly failed: the media and surveys of public opinion. They failed in their capacity to probe important socio-political dynamics and in their predictive capacity, regarding high impact events. These events include the 2016 US presidential election, the Brexit poll and the Colombian 2016 peace agreement referendum. For this reason, new alternatives to measure the political climate have arisen to meet these needs.

Nowadays, the massive use of social networks has allowed for multiple interactions between users, who express their opinions on different topics, people, events and brands. Moreover, the use of social networks in election years, people tend to comment about the candidates, either by their proposals or by their performance in media related events. To extract the relevant information from these political opinions, different Computational Linguistic methods can be applied, such as Sentiment Analysis (SA) on these interactions to get the overall political sentiment.

Twitter is one of the most influential social networks for sharing political messages. This platform is a micro-blogging site, which allows users to broadcast short messages with a maximum of 140 characters (recently increased to 280 characters) called “tweets”. With over 328 million monthly active users and 500 million tweets generated per day [23], Twitter has the potential of becoming a valuable source when analyzing sentiment, and, even more so, political related sentiment in an election year.

During 2017, Chile had three instances of elections (Primaries, First and Second round) providing a rich environment to measure the political climate. Furthermore, given that several studies related to forecasting of elections have been conducted, [6, 8, 12, 20, 27, 31], a similar exercise may be undertaken into the Chilean reality so as to examine the outcome of similar methods. Not only that, but several of these studies have been conducted post hoc, so they cannot be taken as true forecasting.

Given all this, the main goal of this study is to make three predictions Ex Ante of each instance of the 2017 Chilean presidential elections. The approach taken to make these predictions is using Supervised machine learning algorithms with Sentiment Analysis techniques. First a number of experts do a manual pragmatic sentiment labeling over tweets collected over a period of time before the elections, which serves as the input for the different classification models. The tweets then are collected ten days before the prediction day, classified, and the distribution of the overall preferences of those tweets is analyzed to make the prediction. Finally, these predictions are contrasted with the true results of the elections after the event has occurred.

The paper is organized as follows: in Sect. 2, the state of the art is presented. Section 3 describes both the problem and the data used in this study. In Sect. 4, the methodology is shown, detailing the process carried out through the 3 elections, data processing, the metric to be used and the models applied. Finally, in Sect. 5 the results are discussed and in Sect. 6 the conclusions of the study are presented.

2 State of the Art

The development and use of social networks lets millions of users to generate knowledge and, in turn, share it in an easy way that has allowed widespread growth. Given this phenomenon, there is an interest in finding methods to monitor public opinion and behavior, regarding a wide variety of topics. [18, 21, 28]. Such topics include the areas of health, economy, and politics, the latter being the one pertaining to this research.

One way of carrying out this monitoring is by means of Sentiment Analysis, which consists of the use of natural language processing tools, in addition to computational linguistics, in order to assign a polarity value to a document [25]. In the social network context, it has been observed that Twitter may be used as a corpus to which these techniques could be applied, therefore extracting useful information [1, 17, 24].

Regarding the exercise of making political predictions employing social networks, one of the first seminal studies was the one proposed by Tumasjan et al. [31], in which the German federal elections of 2009 were analyzed. In that research it was found that the number of messages analyzed reflected the distribution of the votes in the election. It is worth noting that, albeit this was an initial approach, there are certain studies that thereupon detected particular problems with this method.

Therefore, Gayo-Avello et al. [14, 15] identified certain problems regarding inconsistencies in various studies dealing with predictions carried out using the social network Twitter. These problems ranged from methodological flaws in which the studies were not predictive (post hoc prediction), statistical flaws in which the samples were not representative, and issues related to the training of the models, among other concerns. Furthermore, there are several authors that likewise detected conflictive results and shortcomings in the prediction process using Twitter [10, 16, 20, 22].

With the aforementioned further research, new studies arose, which took into consideration the deficiencies detected by [14]. An example of this is Bermingham and Smeaton [3], where a study applied to the general elections of Ireland as a case study was carried out, integrating sentiment analysis to the prediction process. The authors conclude that Twitter possesses in fact some predictive power, and that it becomes marginally improved when sentiment analysis is incorporated.

Other studies that have used sentiment analysis to make predictions of electoral results have been: the U.K. general elections of 2010 [12], the Dutch senate elections of 2011 [27], the French elections of 2012 [8], the U.K. general elections of 2015 [6] and the U.S. presidential elections of 2016 [29]. All these studies agree that Sentiment Analysis boosts the predictive power of their methods. However, the issue of these studies remains, in their inability to tackle all the problems identified by Gayo-Avello [14]. Regarding this, Beauchamp [2] made a study concerning the extrapolation and interpolation of vote intention in the US presidential elections of 2012 dealing with most of these problems. Nonetheless, it still presents the problematic that is a post hoc prediction, instead of real forecasting.

3 Problem and Dataset

In this section the definition of the problem of this study is introduced: 2017 as an election year for Chile. Also, the dataset and the candidates who participated in each of the elections are described.

3.1 Problem Definition

During 2017 in Chile, there were several presidential referendums for the upcoming presidential period 2018–2022, which were divided up throughout the year in primary elections, first round and second round.

Two political coalitions participated in the primary elections: “Chile Vamos” (Right-wing coalition) and “Frente Amplio” (Left-wing coalition). In the right-wing coalition, there were three participants: Sebastián Piñera, Felipe Kast and Manuel José Ossandón. In the left-wing coalition, there were only two: Beatriz Sánchez and Alberto Mayol. Since there were only two coalitions, this election was taken as two independent elections on the same day, which was July 2nd. The winners of these elections were Sebastian Piñera for “Chile Vamos”, and Beatriz Sánchez for “Frente Amplio”.

In the first round, the two winners of the primary elections participated in an election with six other candidates. These were: Alejandro Navarro, Eduardo Artés, José Antonio Kast, Carolina Goic, Marco Enríquez Ominami and Alejandro Guillier. This election was held on November 19th and the winners of that election were Sebastián Piñera and Alejandro Guillier.

Finally, the second round was carried out on December 17 with the winners mentioned above. The results of this election were that Sebastian Piñera won over Alejandro Guillier with 54.57% of the voting preferences.

Given the sustained growth of social networks in Chile and Latin America [30], these instances presented an interesting test case to do automated SA over the social networks. An immediate application is to analyze the behavior and opinions of the users and their messages on social networks given the participation of the candidates in media events. Although Facebook is the social network with the largest number of interactions, the Twitter API turned out to be more permissive at the time to track the interactions of users. This allows to check derived interactions of media events related to the candidates.

For this particular reason, it is very interesting to track the opinions of the people in presidential election years and find the opinions and preferences of the people regarding the participating candidates. This in order to find indicators/variables that can help in the prediction process. In recent years, traditional instruments (surveys) have failed worldwide to make predictions in different political events [7, 32], Chile in year 2017 being another example of this.

Given all this, the question arises whether the methods based on machine learning using data from social networks serve as a reliable predictor. Conveniently, the nature of this year, allowed to perform three prediction exercises related to this area. Therefore, this study is expected to be valuable for the

body of knowledge related to predictions using social networks, providing an insight into the merits and challenges of the applied approach.

3.2 Dataset

The dataset used for this study corresponds to a compilation of tweets generated during the presidential campaigns of the year 2017 by all the users that made mention of either the presidential candidate's account, or the name and surname of each candidate in the messages. In total, there has been tracking to 11 candidates, from May 14th to December 19th of 2017, being this last date on the day of the Second round of the presidential elections in Chile.

The first thing to mention is that we have gathered two kinds of tweets: the original message and the Retweets (RT). As its name implies, the first consists of a message in which the user wants to express something related to a certain candidate. The number of original messages obtained during each period of time is presented in Table 1.

Table 1. Original tweets collected through the elections periods of the year.

| Corpus | Primary elections | First round | Second round |
|------------------------|-------------------|-------------|--------------|
| Alberto Mayol | 53957 | - | - |
| Felipe Kast | 66914 | - | - |
| Manuel Jose Ossandon | 94564 | - | - |
| Alejandro Navarro | - | 29524 | - |
| Eduardo Artes | - | 10070 | - |
| Jose Antonio Kast | - | 259248 | - |
| Carolina Goic | - | 130196 | - |
| Marco Enriquez Ominami | - | 147941 | - |
| Beatriz Sánchez | 115605 | 197415 | - |
| Alejandro Guillier | - | 34316 | 117258 |
| Sebastian Piñera | 195566 | 572222 | 304911 |
| Total | 526606 | 1380932 | 422169 |

On the other hand, the RT consists of an action by means of which the users are able to replicate an original message as it was written, without adding any content to it. Although this message is the same as the original, it is delivered by another user, providing information for the tweets Sentiment Analysis. The total number of RTs for each candidate during the different election periods are presented in Table 2.

Table 2. Retweets collected through the elections periods of the year.

| Corpus | Primary elections | First round | Second round |
|------------------------|-------------------|-------------|--------------|
| Alberto Mayol | 110514 | - | - |
| Felipe Kast | 138577 | - | - |
| Manuel Jose Ossandon | 140451 | - | - |
| Alejandro Navarro | - | 74143 | - |
| Eduardo Artes | - | 14645 | - |
| Jose Antonio Kast | - | 760073 | - |
| Carolina Goic | - | 355002 | - |
| Marco Enriquez Ominami | - | 368539 | - |
| Beatriz Sánchez | 263900 | 421894 | - |
| Alejandro Guillier | - | 99164 | 344868 |
| Sebastian Piñera | 414321 | 1225873 | 799633 |
| Total | 1067763 | 3319333 | 1144501 |

Regarding both tables, the tracking was limited for each candidate to the round they participated in. One of the features observed is that there is a higher amount of RTs than original messages. This could be relevant, since doing tracking of the RTs, possible influencers of this social network could be detected. Another thing to take into account is that along as the different instances of elections were being conducted, the participation in this social network increased. Finally, the candidate with the highest number of messages was Sebastián Piñera and the one with the lowest number of messages was Eduardo Artes.

As for the manual sentiment analysis of tweets, it was carried out by six experts trained to detect the polarity of the messages. They conducted this labeling process mainly in particular time schedules related to certain media events (interviews, debates, etc.) Concerning the possible sentiments, they correspond to three labels: Positive, Negative, and Neutral. It should be noted that the sentiment analysis approach is based on a pragmatic labeling, rather than a semantic one. This means that the polarity is labeled over the context of the tweet, instead of the semantic polarity of the words composing the tweet. Given this, if a tweet was labeled as positive for a candidate, the feeling is transferred only to it because of the context. Table 3 shows the total volume of tweets tagged for each of the candidates tracked through the three elections.

In the case of labeled tweets, there is no balance between positive and negative classes for each candidate. Regarding neutral tweets, these correspond to most of the tweets labeled for all candidates, with the exception of Manuel Jose Ossandon. Finally, the candidate that generated the most labeled activity was Sebastian Piñera, while the one with least labeled activity was Eduardo Artes.

Table 3. Labeled tweets obtained from the experts from all the elections.

| Corpus | Positive | Negative | Neutral |
|------------------------|----------|----------|---------|
| Alberto Mayol | 5994 | 3711 | 14932 |
| Felipe Kast | 17635 | 5382 | 7383 |
| Manuel Jose Ossandon | 4786 | 13827 | 11908 |
| Alejandro Navarro | 772 | 3581 | 9432 |
| Eduardo Artes | 161 | 2356 | 5570 |
| Jose Antonio Kast | 11327 | 4070 | 55787 |
| Carolina Goic | 7980 | 2901 | 25217 |
| Marco Enriquez Ominami | 15495 | 4421 | 35170 |
| Beatriz Sánchez | 17579 | 25008 | 56648 |
| Alejandro Guillier | 6876 | 9262 | 63925 |
| Sebastian Piñera | 37082 | 51587 | 102459 |
| Total | 125687 | 126106 | 388431 |

4 Methodology

Since throughout the year, three election processes were held, the methodology that was proposed for each of these share certain foundations. The idea behind it is that in order to make the prediction, first a set of tweets is taken before the election and is separated into two parts: Training set and Prediction set.

The first set consists of all the tweets that have a label within a certain date range, which will serve to train a supervised learning algorithm. The prediction set, on the other hand, consists of all the tweets regardless of the label, then all the tweets are selected from the final date of the Training set to the date when the elections will be held. This range of dates is called the prediction window.

Regarding the labels that were used, as the manual classification was carried out with a pragmatic approach in which the sentiment is directed to the candidate, positive labels were used for the training of the algorithm. This is mainly done by transcribing a positive label to “positive-candidate”. With the new labels, the classifier will be trained for all the candidates that participated in that election, with the tweets labeled for the training set. Once the classifiers have been trained, they are applied to the total volume of messages within the prediction window. This allows to make the prediction for this gross amount, obtaining a number of preferences of candidates, which are then converted into the percentages of the prediction.

For the primary elections, a prediction window of 10 days was adopted, and because it was the first predictive exercise, a delta of 3 days before the election was taken. This was mainly done by what is described in [9], where the authors indicate that while daily monitoring of social networks is indeed convenient, there is some evidence that the prediction can be made days before the event.

With the results obtained from the primary elections, a post-election prediction process was carried out. This process had as a goal to be able to adjust the models and obtain some information on how the different prediction windows and training sets behaved. This was done with the aim of applying this knowledge in the following elections.

For the elections of the first round a prediction window of 10 days was adopted, as well as for the primary elections. Using the results obtained with the primary exercise, a date for the prediction 6 days before the election was chosen.

Finally, with the results obtained, for the second round it was decided to monitor daily the political preferences. This was done due to the unsatisfactory results obtained with the election methods of the first round, as it will be detailed in the results section. In this sense, the 10-day prediction window was also kept, but the prediction date was changed to the day before the elections. Table 4 provides a summary of what was described above for all elections.

Table 4. Date range for the training/prediction sets and how many days before the election the prediction was made.

| Election | Training set dates | Prediction set days | Days before election |
|--------------|-----------------------|----------------------------|----------------------|
| Primaries | June 1st–June 19th | June 19th–June 29th | 3 days |
| First round | June 1st–November 6th | November 6th–November 16th | 5 days |
| Second round | June 1st–December 6th | December 6th–December 16th | 1 days |

In the case of the tweets themselves, a preprocessing of the texts was performed. This includes the elimination of stopwords in Spanish, normalization of text, eliminating accent marks, removing scores and symbols. Moreover, the entire messages were taken in lower case. In addition to this, web links and all mentions to Twitter user accounts (username) were removed as well. The latter was done because they can give noise to the automatic training process of the classifiers. Finally, it was also decided to leave the emojis in the tweets (a digital image used to express an emotion or idea.) Hashtags (#word) were also kept, because they can provide useful distinctive information to carry out the classification among candidates.

In order to confirm that the predictions are correct, the results of the elections will be used as ground truth. The metric used to measure this comparison is the Mean Absolute Error (MAE) of the prediction. The MAE is computed as:

$$MAE = \frac{1}{n} \sum_{i_1}^n |y_i - x_i| \quad (1)$$

Where y_i corresponds to the prediction made for the i element, and x_i corresponds to the real i element value (in this case, the percentage of the electorate for a candidate).

In order to carry out the training and prediction, the texts were transformed using a word bag approach and a unigram representation. In the case of the algorithms used, they will be detailed in the following subsection.

4.1 Algorithms

In this subsection the six different baseline models are presented. These are: AdaBoost, Support Vector Machines (SVM), Decision Trees, Random Forest and two voting classifiers. The selection of these models is due to the following: SVM's have been used since the very beginning of SA, and have proved to deliver good performances with different data sets [19]. On the other hand, there are not many studies detailing the performance for political sentiment analysis with AdaBoost and Random Forest, therefore the decision to use them corresponds to the desire to contribute to the general knowledge on how these algorithms perform, in this context.

AdaBoost. AdaBoost (Adaptive Boosting) [13] is an ensemble method that combines weak classifiers which have relatively good classification accuracy, in order to make one strong classifier. This process begins with training N classifiers with modified versions of the data. Subsequently, the individual predictions are merged through a weighted majority vote to make the final prediction. These weights are updated on each iteration of the boosting algorithm, where erroneous instances are weighted up and vice versa. A Decision tree was used as the weak classifiers and the hyperparameter tuned in this study corresponds to the amount of estimators used for the creation of the ensemble.

Support Vector Machines. Support Vector Machines [11] algorithms are a supervised learning method for both classification and regression. SVMs represent the training data as points in a space and use hyperplanes to make separations between classes. Afterwards, by using that space new examples are projected, and the prediction is made according to which side of the separation the projection falls on. To make the hyperplanes, the SVMs implements kernels which allows them to make both linear (linear kernel) and nonlinear separations (polynomial and radial basis function kernels). For this study, a SVM with a linear kernel was implemented, which provides one hyperparameter to tune: cost of misclassification of the data on the training process (C).

Decision Trees. Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. Their aim is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data. The algorithm used in this study was Classification and Regression Tree (CART) [4], which is based on the C4.5 tree algorithm. The hyperparameters that were tuned were the separation criteria and the depth of the tree.

Random Forest. Random forest (RF) [5] is an ensemble method, which uses bootstrap aggregation (bagging) over the original features to build a large number of decision trees. The main objective for RF is to deal with over-fitting and also reduce the variance between trees. The classification then is computed with the mode of the outputs of each tree within the forest. For RF, the hyperparameter is the number of DTs used, depth and separation criteria when building the DTs.

Voting Classifier. Voting Classifier is an ensemble method, which groups several classifiers to do the classification. These are trained in parallel with the same data, and then vote to make the prediction of a sample. In this study, two voting classifiers have been used, these being Hard and Soft voting classifiers. The Hard Voting (V1) classifier makes a classification by majority vote, while the latter (v2) makes the prediction selecting the largest sum of the predicted probabilities for all classifiers. These voting classifiers are built with each method described previously (AdaBoost, DT, SVM and RF).

4.2 Implementation

In this study, the library scikit-learn [26] was used for the implementation of the different algorithms described in the previous subsection. For the primary elections, the machine used had an i5 3.2 processor and 16 GB RAM. For the first and second round elections, the machine used had an i9-7900X processor with 128 GB RAM.

5 Results

As detailed in Sect. 3.1, two coalitions participated in the primary elections, so the prediction exercise was carried out as if they were two separate elections. For this reason, the MAE is reported separately, one for each coalition. The MAE obtained with the different algorithms for all elections is presented in Table 5.

Table 5. Mean Absolute Error for each algorithm in each election.

| Algorithm | P. Chile Vamos | P. Frente Amplio | First round | Second round |
|-----------|----------------|------------------|---------------|---------------|
| AdaBoost | 0.0346 | 0.1684 | 0.0656 | 0.0117 |
| LSVM | 0.1533 | 0.1587 | 0.0909 | 0.1923 |
| DT | 0.1036 | 0.2363 | 0.0737 | 0.1706 |
| RF | 0.0860 | 0.2560 | 0.0648 | 0.1750 |
| V1 | 0.0563 | 0.2878 | 0.0635 | 0.0150 |
| V2 | 0.1042 | 0.2057 | 0.0700 | 0.0051 |

As stated before, the primary elections were divided into two blocks (P. Chile Vamos and P. Frente Amplio). The best results were obtained by AdaBoost with an MAE of 3.46% for Chile Vamos' primary elections. For the Frente Amplio the lowest MAE was obtained LSVM with a 15.87%. Next, in Fig. 1 the percentages predicted with AdaBoost vs the actual results of the election are presented.

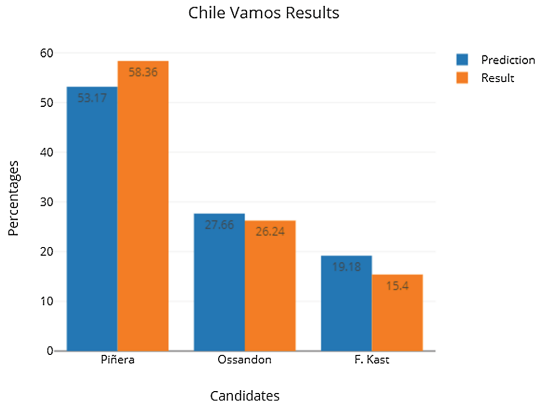


Fig. 1. Percentages of the prediction versus the real values for Chile Vamos election using AdaBoost

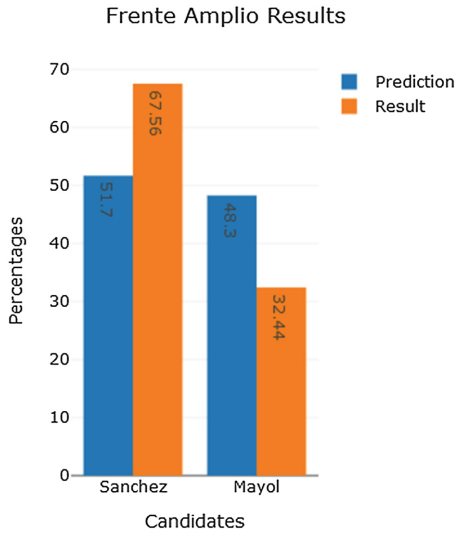


Fig. 2. Percentages of the prediction versus the real values for Frente Amplio election using Linear Support Vector Machine

On the other hand, Fig.2 presents the results obtained for the Frente Amplio’s primary elections, where there is a clear deviation of the prediction versus the real electoral votes.

Regarding the elections of the first round, the Hard voting classifier (V1) obtained the best results, achieving an MAE of 6.35%. Although all the classifiers obtained an MAE under 10%, it was found that when comparing the results of the prediction versus the real values, all the classifiers had a tendency to give a greater favoritism to J. Kast than to Guillier (runner up of the first round).

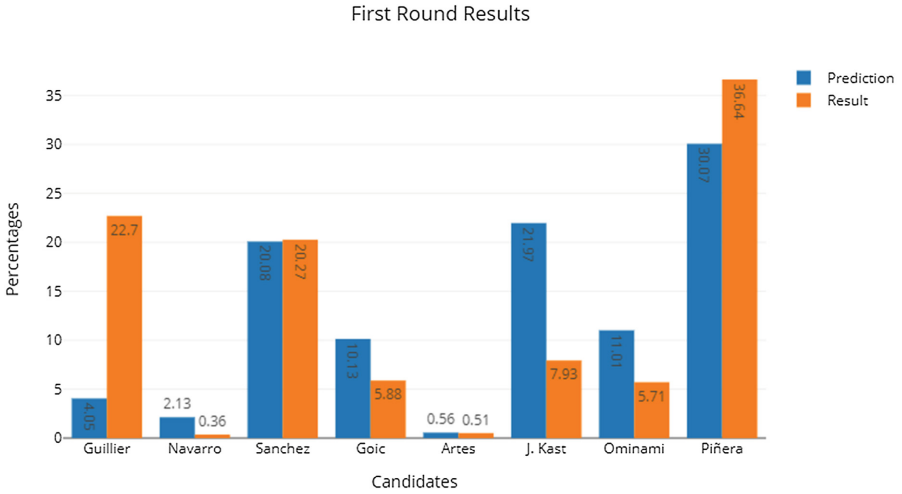


Fig. 3. Percentages of the prediction versus the real values for first round election using Hard Voting Classifier

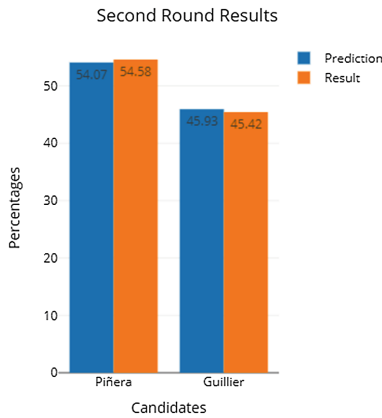


Fig. 4. Percentages of the prediction versus the real values for second round election using Soft Voting Classifier

The Fig. 3 shows what was indicated previously in the prediction obtained by the best classifier for this election.

Finally, for the second round the soft voting classifier (V2) obtained an MAE of 0.51%, being the lowest MAE throughout the study. Figure 4 shows the prediction results for V2, detailing the narrow margin between the prediction and the real election result.

6 Discussion

With the results obtained, AdaBoost had an overall good performance. This is because it obtained the best result in one of the elections (P. Chile Vamos), and results closer to the lowest MAE consistently. It must be pointed out that although each election was made throughout the year 2017, the use of social networks in political campaigns was used in a more active manner in the first and second round. Apart from this, the immediate deployment of these models to obtain early information in the primaries could have worked against the aim of the study. For this reason, the primary elections served as a good prediction exercise, both to get an idea of what the elections were going to be like, and to refine the hyperparameters of the models in order to obtain the lowest MAE possible. The main objective of this was to prepare for the first and second round elections.

Concerning the results of the primary elections, the results obtained were close to the real percentages for Chile Vamos' election, with an MAE of 3.46%. On the other hand, Frente Amplio's results had an MAE of 15.87%, where the prediction was correct on who was going to win (Fig. 2), although the values were further apart from the real electorate percentages. It should also be noted that the best classifier for Frente Amplio's prediction (LSVM), although it obtained the lowest MAE for that prediction, in the other elections its performance was the worst in all cases. This is an important issue for future research.

For the first round although an MAE of 6.35% was obtained, the results of the prediction show a bias for J. Kast. This finding is not surprising, given that he had a strong social network campaign over the last 2 weeks before the elections. Due to this, the bias probably was related to one of the main concerns proposed by Gayo-Avello [14] related to "all the tweets are assumed to be trustworthy". Regarding this, Post hoc analysis showed that malicious behavior related to false accounts generating fake activity in favor of the Candidate (Astroturfing), and that it was present not only for J. Kast, but other candidates as well.

Regarding these same elections, the reason behind the bad performance of candidate Guillier (runner up), is due to the fact that there was an error in capturing the interactions for said candidate. This candidate had two official accounts and apart from this, the name "Alejandro Guillier" was tracked. The problem with this is that the users on twitter misspelled his name as "Guiller"; hence this was possibly the main source of the problems of the prediction for this candidate. This tracking error was detected and fixed in the previous days of the election day. This can be seen in the total of tweets for the candidate in

the first round versus the second round (Tables 1 and 2). Finally, this increase may also be due to the fact that the different left-wing candidates urged their voters to give their support to Guillier, potentially increasing the discussion on twitter.

On the other hand, the results obtained in the second round were close to the electoral results, with an MAE of 0.51% for the Soft Voting Classifier (V2). These results are attributed to the amendment for tracking the candidate Guillier.

Regarding the raw volume of tweets prediction discussed in [31], low MAE results were obtained for the primaries. However, when this method was applied to the other two elections, the prediction distances from the true values as detailed in Table 6.

Table 6. Mean Absolute Error for raw volumes of tweets analysis.

| P. Chile Vamos | P. Frente Amplio | First round | Second round |
|----------------|------------------|-------------|--------------|
| 0.0278 | 0.0221 | 0.0673 | 0.1765 |

Finally, the lack of a description of the demography of the users, and the approach that each tweet of the prediction set is a vote towards a candidate after the classification process are issues also detailed by Gayo-Avello. This could explain easily the high values of MAE obtained in the Frente Amplio’s primary elections, or the predicted value for J. Kast for the first round. Taking the latter as an example, it could be seen that J. Kast had a very strong political campaign at the end of the last two weeks before the election as mentioned before. This increases the number of tweets that express a positive sentiment towards the candidate. For that reason, it is urgent to look for a better method to model the vote intention for the Twitter users.

7 Conclusions

In this study three electoral predictions were made through several months of the year 2017 for the Chilean presidential elections. Supervised learning algorithms were trained with pragmatic sentiment labeled tweets and predicted the distribution over a prediction set. Hyperparameters tuning for both algorithms and training/prediction set were conducted from the primaries election and first round, resulting in a final accurate prediction with an MAE of 0.51% with the Soft voting classifier.

One of the main motivations of this study was to make Ex Ante predictions of the elections, which resulted in a challenging problem. Example of this was to discover after the first round the error of tracking towards the account of the candidate Guillier, or the failed estimate towards the candidate J. Kast giving him the second majority. Mostly, this could be one of the flaws detailed by Gayo-Avello cite Gayo1, which mentions that “all the tweets are trustworthy”.

This leads to the fact that in the event of making a prediction of this style, factors such as astroturfing and the use of social bots must be taken into account; as well as the need to make a review of the demographics of users.

As future work, the use of other labels is proposed to improve the performance of predictive models. As such, this presents a source of information not exploited in the present study. Other relevant information that can be obtained from the total tagged corpus, is the use of topic models both to be able to see the political discourse and the opinions of the Twitter users changing through the electoral period. In addition to this, it would be interesting to use the topic model words obtained and assigning them a greater weight when making predictions.

Acknowledgments. This work was supported by the “Proyectos Interdisciplinarios” Grant of VREIA - Pontificia Universidad Católica de Valparaíso. Héctor Allende-Cid’s work was supported by the “Fondecyt Initiation into Research 11150248” of Conicyt, Chile.

References

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of Twitter data. In: Proceedings of the Workshop on Languages in Social Media, pp. 30–38. Association for Computational Linguistics (2011)
2. Beauchamp, N.: Predicting and interpolating state-level polls using Twitter textual data. *Am. J. Polit. Sci.* **61**(2), 490–503 (2017)
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 (SAAIP 2011), pp. 2–10 (2011)
4. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Chapman & Hall, New York (1984)
5. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
6. Burnap, P., Gibson, R., Sloan, L., Southern, R., Williams, M.: 140 characters to victory?: Using Twitter to predict the UK 2015 general election. *Elect. Stud.* **41**, 230–233 (2016)
7. By, R.T., Ford, J., Randall, J.: Changing times: what organizations can learn from brexit and the 2016 us presidential election. *J. Change Manag.* **17**(1), 1–8 (2017). <https://doi.org/10.1080/14697017.2017.1279824>
8. Ceron, A., Curini, L., Iacus, S.M., Porro, G.: Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens’ political preferences with an application to Italy and France. *New Media Soc.* **16**(2), 340–358 (2014). <https://doi.org/10.1177/1461444813480466>
9. Ceron, A., Curini, L., Iacus, S.M.: Politics and Big Data: Nowcasting and Forecasting Elections with Social Media. Taylor & Francis, New York (2016)
10. Chung, J., Mustafaraj, E.: Can collective sentiment expressed on twitter predict political elections? In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, pp. 1770–1771, AAAI 2011. AAAI Press (2011). <http://dl.acm.org/citation.cfm?id=2900423.2900687>
11. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)

12. Franch, F.: (Wisdom of the crowds)2: 2010 UK election prediction with social media. *J. Inf. Technol. Polit.* **10**(1), 57–71 (2013). <https://doi.org/10.1080/19331681.2012.705080>
13. Freund, Y., Schapire, R.E.: A short introduction to boosting. In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 1401–1406. Morgan Kaufmann, San Francisco (1999)
14. Gayo-Avello, D.: No, you cannot predict elections with Twitter. *IEEE Internet Comput.* **16**(6), 91–94 (2012)
15. Gayo Avello, D., Metaxas, P.T., Mustafaraj, E.: Limits of electoral predictions using twitter. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence* (2011)
16. Jungherr, A., Jürgens, P., Schoen, H.: Why the pirate party won the German election of 2009 or the trouble with predictions: a response to Tumasjan, A., Sprenger, T.O., Sander, P.G., & Welpe, I.M. "predicting elections with Twitter: what 140 characters reveal about political sentiment". *Soc. Sci. Comput. Rev.* **30**(2), 229–234 (2012). <https://doi.org/10.1177/0894439311404119>
17. Kouloumpis, E., Wilson, T., Moore, J.D.: Twitter sentiment analysis: the good the bad and the OMG!. *ICWSM* **11**(538–541), 164 (2011)
18. Lariscy, R.W., Avery, E.J., Sweetser, K.D., Howes, P.: Monitoring public opinion in cyberspace: how corporate public relations is facing the challenge. *Public Relat. J.* **3**(4), 1–17 (2009)
19. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In: Aggarwal, C., Zhai, C. (eds.) *Mining Text Data*, pp. 415–463. Springer, Boston (2012). https://doi.org/10.1007/978-1-4614-3223-4_13
20. Lui, C., Metaxas, P.T., Mustafaraj, E.: On the predictability of the us elections through search volume activity (2011)
21. Madge, C., Meek, J., Wellens, J., Hooley, T.: Facebook, social integration and informal learning at university: 'it is more for socialising and talking to friends about work than for actually doing work'. *Learn. Media Technol.* **34**(2), 141–155 (2009)
22. Metaxas, P.T., Mustafaraj, E., Gayo-Avello, D.: How (not) to predict elections. In: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pp. 165–171. IEEE (2011)
23. Omnicore Agency: Twitter by the numbers: stats, demographics & fun facts. <https://www.omnicoreagency.com/twitter-statistics/>. Accessed 28 Dec 2017
24. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: *LREc*, vol. 10 (2010)
25. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. *Foundations and trends®. Inf. Retrieval.* **2**(1–2), 1–135 (2008)
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
27. Sang, E.T.K., Bos, J.: Predicting the 2011 Dutch senate election results with Twitter. In: *Proceedings of the Workshop on Semantic Analysis in Social Media*, pp. 53–60. Association for Computational Linguistics, Stroudsburg, PA, USA (2012). <http://dl.acm.org/citation.cfm?id=2389969.2389976>
28. Shirky, C.: The political power of social media: technology, the public sphere, and political change. *Foreign Aff.* **90**(1), 28–41 (2011)

29. Singh, P., Sawhney, R.S., Kahlon, K.S.: Forecasting the 2016 US presidential elections using sentiment analysis. In: Kar, A.K., Ilavarasan, P.V., Gupta, M.P., Dwivedi, Y.K., Mäntymäki, M., Janssen, M., Simintiras, A., Al-Sharhan, S. (eds.) I3E 2017. LNCS, vol. 10595, pp. 412–423. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68557-1_36
30. Fung Global Retail & Technology: Deep dive social media in Latin America. Technical report, May 2016. <https://www.fbicgroup.com/sites/default/files/Social>
31. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with Twitter: what 140 characters reveal about political sentiment. *ICWSM* **10**(1), 178–185 (2010)
32. Valentino, N.A., King, J.L., Hill, W.W.: Polling and prediction in the 2016 presidential election. *Computer* **50**(5), 110–115 (2017)