USING SOCIAL MEDIA TO FORECAST ELECTORAL RESULTS: A REVIEW OF THE STATE OF THE ART

Andrea Ceron[1], Luigi Curini

Department of Social and Political Sciences, University of Milan, Italy

Stefano Maria Iacus

Department of Economics, Management, and Quantitative Methods, University of Milan, Italy

Received: 3rd April 2014/ Accepted: 30th November 2015

Abstract. *The paper discusses the advantages of using Supervised Aggregated Sentiment Analysis (SASA) of social media to forecast electoral results and presents an extension of the ReadMe method (Hopkins and King, 2010), which is particularly suitable to addressing a large number of categories (e.g. parties) providing lower standard errors. We analyze the voting intention of social media users in several elections held between 2011 and 2013 in France, Italy, and the United States. We then compare 80 electoral forecasts made using these or other techniques of data-mining and sentiment analysis. The comparison shows that the choice of the method is crucial. Electoral forecasts are also more accurate in countries with higher Internet penetration and given the presence of electoral systems based on proportional representation.*

Keywords: *Electoral Forecast, Sentiment Analysis, Text Mining, Social Media, Twitter.*

1. INTRODUCTION

The “big data revolution is providing scholars with highly informative data concerning several areas of study and potentially yielding monumental consequences in the real world (King, 2014). The diffusion of the Internet as well the striking growth in social media and social network sites, such as Facebook and Twitter, represent one of the primary sources of such a revolution. As millions of citizens start to surf the web, creating their own account profiles and sharing information on-line, a substantial amount of data becomes available. This information can be successfully exploited to explain and anticipate dynamics concerning different

[1] Andrea Ceron, email: andrea.ceron@unimi.it

topics, such as stock markets, movie success, and disease outbreaks (Ceron, Curini and Iacus, 2013; Kalampokis, Tambouris and Tarabanis, 2013; Schoen et al., 2013).

The Internet is also playing a role in favouring political mobilization (Bennett and Segerberg, 2011; Cottle, 2011; De Zuniga et al., 2009; Larsson and Moe, 2012) and provides a “public sphere where citizens and politicians can broadcast their own political views (Papacharissi, 2002). Accordingly, the concept of exploring social media to analyze the political preferences of individuals (O’Connor et al., 2010; Woodly, 2007) as well as to nowcast electoral campaigns (Ceron, Curini and Iacus, 2013, 2015; Ceron and d’Adda, 2015) and (eventually) forecast electoral results seems to flow naturally from such developments (Ceron et al., 2014; Gayo-Avello, 2012, 2013).

Considering that the names of front-runner candidates are widely known and that people talk about them, earlier attempts focused on quantitative data and measured the volume of conversations related to rival politicians (Veronis, 2007) or the number of “likes received on Facebook pages (Giglietto, 2012) to predict the share of votes for each candidate. In addition, voters may directly report the name of the candidate they plan to support. Based on this, Tumasjan et al. (2011) disclosed the predictive power of social media and claimed to have predicted the 2009 German Federal Elections, reporting, on average, a 1.65 point deviation between the share of Twitter mentions and the actual share of votes won by the main German parties.

However, the number of mentions received by a party/candidate, *per se*, is not necessarily informative about the support enjoyed by the same party/candidate within the population of (Internet) voters. Candidates can be mentioned not only by those who are willing to vote for them but also by voters who want to criticize them. For these reasons, other studies have adopted techniques of sentiment analysis (SA) mainly based on ontological dictionaries to supplement the analyses of the volume of data.[2]

For instance, Lindsay (2008) developed a sentiment classifier based on lexical induction and reported correlations between polls and the content of wall posts available on Facebook during the 2008 U.S. presidential election. Similarly, O’Connor et al. (2010) confirmed that Obama’s approval rating is correlated with the sentiment expressed by Twitter users. Furthermore, SA performed as well as polling in predicting the results of the 2011 Dutch Senate election (Tjong Kim Sang and Bos, 2012), while analyses of multiple social media sites (Facebook,

---

[2] Sentiment analysis consists of analyzing texts to extract information.

Twitter, Google, and YouTube) outperformed traditional surveys in estimating the results of the 2010 United Kingdom election (Franch, 2013).

Attempts to forecast elections that adopt one of the two above-mentioned techniques or a combination of the two cover a huge sample of countries, such as Ireland (Bermingham and Smeaton, 2011), Portugal (Fonseca, 2011), Singapore (Choy et al., 2011; Skoric et al., 2012), the Netherlands (Sanders and den Bosch, 2013), the United Kingdom (Lampos, 2012; Tweetminster, 2010); and the United States (Choy et al., 2012; Di Grazia et al., 2013; Jensen and Anstead, 2013; Shi et al., 2012; Washington et al., 2013).

These pioneering analyses, however, are not exempt from criticism (Gayo-Avello, 2012; Jungherr et al., 2012; Metaxas, Mustafaraj and Gayo-Avello, 2011). The main criticisms concern the fact that naïve SA techniques do not consider humour, double meanings and sarcasm and therefore may not accurately assess the actual meaning of the opinions expressed. In addition, automated/unsupervised SA software cannot address the risk of a spamming effect due to the presence of rumours and misleading information. Further, counting the number of positive and negative terms in a sentence may lead to paradoxical effects (think about this sentence from a movie review: "a *great* cast, *good* visual effects, the actors give their *best*, the script was *promising*, but this movie *sucks*").

Such concerns, however, can be addressed by relying on a proper method for sentiment analysis. Recent developments in quantitative text analyses allow one to *integrate* quantitative large-n data with in-depth analyses and re-open the debate around whether analyses of social media can be helpful in predicting election results.

In this paper, we present a modified version of the technique of Supervised Aggregated Sentiment Analysis (SASA) known as ReadMe, which was originally proposed by Hopkins and King (2010). The modification improves the ability of the standard ReadMe approach in discerning between *noise* and *signal* via constrained estimation, exploiting the fact that ReadMe focuses on the estimation of the aggregated distribution of opinions rather than on the individual classification of each single text. As a by-product of this enhancement, the forecasting power of sentiment analysis over social media data is vastly improved as well. We call this method *i*SA (*integrated* SA). We apply both ReadMe and this new method to forecast several elections held in France, Italy and the United States between 2011 and 2013. We then contrast our results with those reported by other techniques of data-mining or sentiment analysis by analyzing 80 social-media-based electoral forecasts. In this way, we show that SASA, and in particular *i*SA, dramatically increases forecast accuracy. The analysis also reveals that, overall, social media data are better predictors of election outcomes when the share of Internet users within a country is relatively higher and given the presence of electoral systems either based on proportional representation or in which voters cast a vote for a specific candidate to a monocratic position rather than choosing, for example, from a party list.

The next section will briefly present the ReadMe approach before discussing the novelties introduced by the *i*SA method and comparing both methods of SASA with traditionally employed alternatives. The following sections are then devoted to analyzing the accuracy of the electoral forecasts based on the analysis of social media.

2. SUPERVISED AGGREGATED SENTIMENT ANALYSIS: README AND ISA

The ReadMe approach, which represents the first example of what we call SASA, was introduced by Hopkins and King (2010) to solve two different problems that affect all other methods of sentiment analysis when they are applied, in particular, to data originating from social media. The first problem is that users on social media employ natural language, which evolves continuously and varies depending on the person who is actually writing (male, female, young, older, officer, journalist, etc.) and the topic (football, politics, music, etc.). Further, metaphoric or ironic sentences, as well as jargon, contractions or neologisms, are used in different and new ways every time. This fact makes all unsupervised methods based on ontological dictionaries or statistical methods based on NLP (natural language processing) models unable to accurately capture sentiment (by sentiment, here, we refer to any type of semantic meaning, e.g., expression of vote or attitude toward buying a given good). Think, for example, about the informal expression "what a nice rip-off!", which is ambiguous from the viewpoint of an ontological dictionary because it includes both a positive and negative term but would be correctly assigned a negative sentiment by a human coder. For these reasons, supervised human coding of a *training set* is a fundamental step of the SASA method.[3]

---

[3] Just to report a few examples: in the hand-coding stage, we were able to isolate ironic viral phenomena such as the one that occurred in the 2012 primary election of the Italian centre-left coalition, whereby a group called "*Marxisti per Tabacci*" (Marxists for Tabacci) wrote fake statements to support (ironically) one of the candidates (the moderate Tabacci): "*Comrades, I support the motion 'Marxisti per Tabacci'. #primarie @pdnetwork*". In addition, the hand-coding period also allows one to evaluate when a voter explicitly expresses his or her intention to behave strategically, declaring a strategic vote in favor of one of the main candidates though signaling a true preferences for a minor candidate. The following are examples related to the Italian 2013 election and to the U.S. 2012 Presidential election: "*My heart beats for Vendola, but I will vote for Bersani*" and "*I lean 90% with libertarians but I think I'll vote for Obama*".

The second issue is *noise*. Contrary to, for example, official political speech in the Parliament (which makes use of a given vocabulary), most of the time the analysis of social media is performed using web-(or Twitter/Facebook/forum-)crawling based on keywords. Regardless of how they are cleaned (again, typically using unsupervised methods), the data still contain a substantial amount of *noise* (i.e., text that makes use of sentences or words that sound similar to but are unrelated to the topic of interest) or *off-topic* texts. Even if one creates a *training set* using supervised human coding (i.e., without classification errors), the resulting statistical classifier will predict a sentiment for the texts in the actual *test set* that is, with high probability, a sentiment that falls in the off-topic category even when this is not the case. As a result, the aggregation of these individually predicted classifications produces biased estimates of the aggregated distribution as well as estimates with large variability. The aggregated estimation method of the SASA approach directly estimates the aggregated distribution without the need to use individual classifications of texts in the *test set*. In social science as well as in electoral studies, what matters in forecasting is the aggregated distribution of opinion or the share of votes rather than individual opinions or voting behaviour.

The ReadMe method proposed by Hopkins and King (2010) is based on a two-stage process. The *first stage* involves human coders and consists in reading and coding a subsample of the documents downloaded from the Internet/Twitter/etc. This subsample – with no particular statistical properties (see below) - represents a *training set* that will be used by the second step of the algorithm to classify all the unread documents (the *test set*). Human coders are, of course, more effective and careful than ontological dictionaries in recognizing all the previously discussed language specificity issues and the author's attitude on the subject, including those regarding the expression of an opinion that could be eventually linked to a "vote" (see the next section). Specifically, human coding enables a better interpretation of texts by reducing classification errors. Moreover, human coding is better suited for identifying the (ever-present) problem of spamming in social communication as well as tagging off-topic texts. This is, of course, important given that spamming and unrelated content can have a negative impact on the accuracy of the final result.

At the *second stage*, the aggregated statistical estimation of the ReadMe algorithm extends such accuracy to the whole population of posts, allowing one to properly obtain the opinions expressed on social networks. More formally, the ReadMe approach can be described as follows: each text in the data set is decomposed into single words, or *stems*. Each text is than represented as a profile of zeroes and ones. This is called the stemming phase. Stemming can occur in many ways, e.g., by removing all stop words, conjunctions, etc., or words that are always present in all texts. A *stem* can be a single word, a couple or a triplet of words, etc. Words are usually shortened to their root by, for example, removing prefixes and suffixes. However, in general, the order of the words or the structure of the sentence is not maintained in this approach. This is called “bag of words” stemming. Our experience shows that this is more than sufficient, and the general principle is that the more structure that is put in the text, the more the final estimate will be driven by this a priori knowledge (see also Grimmer and Stuart, 2013). For this reason, simple stemming (no pairs or triplets) that leads to a naive bag of words is the most profitable method of preparing the data for the subsequent analysis.

Let U.S. denote with S the word profiles used in the texts and with D people’ opinions expressed in the texts. Let M be the number of stems kept in the stemming phase. The target of estimation is P(D), i.e., the frequency distribution of the opinions over the posting population, a K x 1 vector.

The standard statistical approach is to decompose P(D) in the following way:

$$P(D) = P(D \mid S)\, P(S). \tag{1}$$

P(S), a vector of dimension 2^Mx1, corresponds to a tabulation of frequencies of word profiles in the whole population of texts. P(D|S), a K$\times 2^M$ matrix, is estimated from the *training set* after supervised human coding as $P_T(D|S)$, i.e., the conditional distribution of word profiles within the *training set*.

The standard statistical approach is to use any classifier (e.g., multinomial regression, classification trees, random forests, or support vector machines) to produce individual classifications of posts in the *test set* (i.e., posts belonging to the corpus of texts but not to the *training set*), and category D_i is assigned to each text with some probability, i.e., for a text j in the test set, with word profile S_j, its category is estimated through $P_T(D_i|S_j)$ for i = 1, 2, …, k. Then, the aggregated distribution of opinions P(D) of all texts in the corpus is obtained by aggregating individual classifications, each with its own classification error. Unfortunately, due to the noise in the social media texts, the estimates of $P_T(D_i|S_j)$ are biased and exhibit high variability because, within the subsample of text with sequence S_j, the frequency of text that belongs to the category D_k=“off topic/noise” is almost one, and all the other categories D_j, $j \neq k$, are very close to zero.

As a result, the individual classification error is high and does not vanish due to aggregation because of the large variance in estimates; in contrast, it easily propagates up to the extent that, in many applications with thousands or millions of texts, one could see the error increasing to 15-20%. This is clearly *quite* problematic if one is mainly interested in estimating some type of aggregate measure through the analysis of social media, as occurs precisely with all the studies that want to map (written) opinions to votes.

The ReadMe method reverses the approach, and instead of estimating the individual opinions and subsequently aggregating them, it directly estimates the aggregated distribution of opinions by aggregating all the word profiles, leading to an error on the order of 2-3%. Formally, one can see (1) in the following different manner:

$$P(S) = P(S|D)\ P(D). \tag{2}$$

The frequency distribution P(S) can be evaluated by tabulating all the texts in the database, which requires only some computer time and no debatable assumptions. The conditional distribution P(S|D), a 2^MxK matrix, cannot be observed on the whole data set and must be estimated by hand-coding of a *training set* of texts, as before.

The hand coding of the training text allows one to calculate $P_T(S|D)$, i.e., the conditional frequency distribution of word profiles inside the *training set*. The assumption – and the reasonable requirement – of the method is that the texts of the training set are homogeneous to the whole data set, i.e., they come from the same "world" as the remainder of the dataset; thus, one can assume that

$$P_T(S|D) = P(S|D). \tag{3}$$

If this is the case, the frequency distribution of the opinions can be consistently estimated because both P(S) and $P_T(S|D)$ are observable. Therefore, by equation (2) and by noticing that $P_T(S|D)$ and P(S|D) are both matrixes, we have

$$P(D) = P(S|D)^{-1}\ P(S) \tag{4}$$

and the final estimate is obtained by replacing P(S|D) with $P_T(S|D)$

$$P(D) = P_T(S|D)^{-1}\ P(S). \tag{5}$$

where $P_T(S|D)^{-1}$ is the inverse matrix of $P_T(S|D)$ and similarly for $P(S|D)^{-1}$, which is the inverse matrix of P(S|D)[4]. It is worth remarking that – while the homogeneity of the training set to the test set is required – no statistical property must be satisfied by the former: in particular, the training set is not a representative sample of the population of texts. The only constraint is that for each category D, a sufficient amount of text (the authors empirically set this number to 20-50 entries) has to be hand coded; thus, the selection of the sample is random but sequential.

Table 1 summarizes the methods available in the literature for analyzing social media (as well as any other type of text in digital format; see Grimmer and Stewart,

---

4 On the technique to invert this matrix see Hopkins and King (2010).

2013) according to the two criteria discussed above: first, the method employed to classify texts (unsupervised versus supervised methods) and, second, the method adopted to estimate the overall distribution of opinions in the classified texts (aggregation of the individual classification of all posts versus direct measurement of the aggregate distribution). Traditionally, studies on social media and elections have focused on the upper-left cell by counting the number of mentions related to a candidate, for example, or by employing ontological dictionaries. The SASA approach in ReadMe, in contrast, focuses on the lower-right cell. The remaining two cells of the table have been up to now scarcely visited by scholars.

Table 1: A Typology for Classifying Texts Posted on Social Network Sites Method to estimate the distribution of opinions

		Individual	Aggregate
Method to classify texts	**Unsupervised**	Counting mentions, ontological dictionaries	=
	Supervised	=	SASA (ReadMe, *i*SA)

In summary, with a specific reference to electoral forecasts, the SASA approach presents two crucial advantages compared to traditionally employed alternatives: first, it enables a better interpretation of the texts during the human-coding stage (as we will see more precisely in the next section) as well as a better treatment of off-topic posts; second, it produces more reliable aggregate results in a context where the aggregate results (i.e., the final vote-share of parties and/or candidates) are what concerns U.S..

ReadMe still fails when the number of categories D is too high for both numerical (the matrix dimensions do matter) and statistical reasons (one needs to classify a larger *training set* to obtain a sufficient amount of coding for each category). This is sometimes the case in elections where the opinions correspond to the number of parties. When parties gather in coalitions, it is much easier to obtain good estimates of the shares for a coalition than of the shares for each individual party.

The *i*SA approach, which represents in several aspects an extension of ReadMe, exploits this by constraining estimation in the following way. Step 1: the shares of the coalitions are estimated via the *i*SA algorithm (see below), using as the target only the coalitions and the off-topic category (for example, 3 to 5 categories). This gives precise estimates of the coalitions' shares. Step 2: each text that falls in a given coalition is recoded according to the intention to vote for an individual party. Only the subset of coded text belonging to the coalition will be reclassified using the *i*SA algorithm, again producing a constrained distribution of voter share within the coalition. This provides precise estimates of vote shares for each party of the coalition, as well. Step 2 is then replicated for all the coalitions, resulting in a two-stage constrained, supervised, and aggregated sentiment analysis. Accordingly, *i*SA is particularly well-suited when there is the need to reduce the number of categories, e.g., coalitions rather than parties.[5]

More precisely, assume a political system with C possible coalitions c_i, $i = 1,\ldots C$. Each coalition c_i consists of several parties; for example, p_{i1} is the first party in coalition i. In Step 1, the texts are classified according to coalitions (suppose that we have four coalitions, i.e., $C = 4$), and the proportion of shares of each coalition is estimated using SASA. Suppose that we obtain $c_1 = 45\%$, $c_2 = 25\%$, $c_3 = 20\%$, and $c_4 = 10\%$. Each text in the training set *tagged* for an individual coalition is hand coded again (this second manual tagging occurs during the coding in Step 1) according to the parties of that coalition plus the additional category, "other coalition". Assume that for coalition c_1, we have three parties. After recoding and SASA analysis, we obtain the following estimates: $p_{11} = 75\%$, $p_{12} = 15\%$, and $p_{13} = 10\%$. In Step 2, we renormalize these estimates of p_{11}, p_{12} and p_{13} to sum to the share of coalition c_1. Therefore, we set $p'_{11} = p_{11} / c_1 \times 100\%$, obtaining $p'_{11} = 33{,}75\%$, $p'_{12} = 6{,}75\%$, and $p'_{13} = 4{,}5\%$. Notice that in this fictitious example, the estimates of p_{13} have been constrained, reducing the bias in the estimation of $P(S=s|D=p_{13})$ due to the small amount of texts expressing that opinion in the training set. In addition, notice that given the faint signal for party p_{13} in the whole data set, the direct estimate of p_{13} combined with all the parties (approximately 18-20) would have led to not only bias but also to large standard errors.

From a technical point of view, regardless of the two-stage approach, *i*SA produces more accurate estimates than the ReadMe algorithm: while the latter estimates P(D) through equation (5) using a simulation approach, *i*SA directly evaluates the estimates. In estimating (5), ReadMe proceeds by randomly sampling the columns of the stem matrix (i.e., randomly selecting the stems), thus reducing

---

[5] For example, in the 2013 Italian political elections, there were more than 20 parties organized into 3 coalitions, plus additional parties, which made the statistical problem quite challenging. Using the constrained estimation approach of iSA, however, it was still possible to predict fairly well the final electoral outcome, as we will see below.

the computational complexity in order to estimate $P_T(S|D)$ and invert the matrix in (5), estimates P(D) over several replications, and averages the estimates of P(D) to produce the final estimate in (5). The algorithm at the base of *i*SA considers all the stems at once and, by collapsing the matrix of stems into a sequence of one-dimensional strings, it uses optimized linear algebra inversion algorithms to invert the corresponding full form of $P_T(S|D)$. In this case, standard errors of the estimates can be obtained with direct calculation without relying on bootstrap methods (as is the case for the ReadMe approach).

More precisely, while ReadMe uses a bagging approach by sampling the stems in order to reduce the dimensionality of the stem matrix and apply a quadratic optimization algorithm to the reduced matrix, *i*SA collapses the space into a one-dimensional quantity and solves the problem in one step. Both methods are asymptotically unbiased, but *i*SA is more efficient in the sense that its variance is smaller than that in the corresponding ReadMe solution, in that *i*SA is an exact method, unless the number of replications in the bagging step is increased to a point that makes the computational times of ReadMe unrealistic. *i*SA is also faster and can be applied to real-time analyses, although this is not the focus of this paper. Furthermore, while convergence of the ReadMe method requires a substantial number of iterations, it becomes unstable when the number of categories D grows for a given sample size, i.e., sometimes the algorithm does not converge if the dimension of D is large or if there are insufficient hand-coded texts for a given category "D = d". However, the *i*SA method is not affected by the dimensionality of the stem matrix, the dimension of the dataset (the extent of the corpus of texts) and the number of categories D. The precise numerical and statistical description of iSA, though, is a matter for a different paper.

In summary, ReadMe and *i*SA represent two possible variants of the SASA methodology that share essentially the same idea as well as the two-stage process. In this paper, *i*SA and ReadMe are not explicitly contrasted in the empirical analysis; therefore, only the performance of the method used at the time of the analysis has been reported (see below).

3. ELECTION FORECASTING THROUGH SASA

By adopting the SASA approach described above, we systematically measured the voting intentions expressed on social media in several elections held in France, Italy and the United States between 2011 and 2013. In this respect, P(D) refers to the (aggregate) propensity to vote for each candidate/party. The peculiarity of SASA is that it allows one to measure the sentiment beyond the concept of positive/

negative statements. We can determine the unsolicited "voting intentions" freely expressed on-line.

To this aim, we deem that a post expresses a real intention to cast a vote in favour of a candidate/party only if at least one of the following three conditions is satisfied: a) the post includes an explicit statement related to the willingness to vote for a candidate/party; b) the post includes a statement in favour of a candidate/party together with a message or a hashtag connected to the electoral campaign of that candidate/party; or c) the post includes a negative statement opposing a candidate/party together with a message or a hashtag connected to the electoral campaign of a rival candidate/party.

Considering a positive statement plus a campaign message or a hashtag, and not simply a generic positive statement, permits one to focus on signals that are more "costly" in terms of self-exposition and that are therefore more credible (for this point, see the sizeable literature on signalling games: Banks, 1991). In contrast, condition c) allows one to reduce the arbitrariness in the "supervised" stage of the analysis. This also applies to a (largely) two-candidate case, such as the U.S. Presidential race. For example, if a tweet says "do not vote for Romney", this does not necessarily imply that the person who wrote that post will then vote for Obama. He or she could decide to vote for a third candidate or to abstain. In a multi-party race, of course, this problem is even more significant. Returning to the previous example, a hypothetical tweet such as "do not vote for Romney. #fourmoreyears" would be counted, according to our classification, as a vote in favour of Obama, given that #fourmoreyears has been one of most largely used hashtags supporting the Obama's electoral campaign.

In Table 2, we list the analyzed elections. Overall, more than 110 million comments have been analyzed.[6] For each election, the Mean Absolute Error (MAE) of the prediction has been reported. It is measured as the absolute difference between the actual results and the forecast and is averaged over the number of parties/candidates considered in the analysis. The MAE has been widely used to compare the accu-racy of forecasts based on social network analyses (e.g., see

[6] The population of tweets (or blogs) that we collected consists of all the tweets posted during the pertinent time period that include in their text at least one of a set of keywords (the name of the political leaders/parties covered by each of our analysis as well as the most popular hashtags characterizing each candidate/party's electoral campaign). Duplicate tweets have been removed. The data have been downloaded through the social media monitoring engine developed at Voices from the Blogs (http://voicesfromtheblogs.com/), while the analysis has been performed using R.

Tumasjan et al., 2011), and we follow this strategy here. With the Milan municipal election of 2011 as the only exception, which has been analyzed ex post focusing on blogs, all the other predictions have been carried out using Twitter data, and the results have been made publicly available *before* the elections. In this sense, they can be strictly considered as real forecasts. With respect to the particular types of SASA approaches discussed in the previous section, we adopted both ReadMe (in our first attempts of electoral forecasts) as well as *i*SA (in the later ones). Whenever available, we also provide a comparison with the estimates produced by other analysts adopting naïve sentiment analysis or counting the number of mentions on social media, which is broadly defined (either using Twitter, Facebook, blogs, or all of them). When multiple forecasts were given, we reported their average value.

Table 2: List of predictions based on the SASA method and comparison with predictions made through alternative techniques (type of SASA approach employed in brackets)

Election	Mean Absolute Error (MAE)	
	SASA	**Mentions/Naïve SA**
Milan municipal election 2011 (second round)	2.60 (ReadMe)	n.a.
French presidential election 2012 (first round)	4.65 (ReadMe)	5.04
French presidential election 2012 (second round)	3.30 (ReadMe)	n.a.
French legislative election 2012 (first round)	2.38 (ReadMe)	n.a.
U.S. presidential election 2012 (popular vote)	0.02 (ReadMe)	6.85
Centre-left primary election 2012 (Italy) (first round)	1.96 (*i*SA)	8.59
Centre-left primary election 2012 (Italy) (second round)	1.50 (*i*SA)	n.a.
Italian general election 2013	1.62 (*i*SA)	7.68
Lombardy regional election 2013	1.59 (*i*SA)	n.a.
Democratic Party primary election 2013 (Italy)	9.17 (*i*SA)	8.51
Northern League primary election 2013 (Italy)	0.50 (*i*SA)	n.a.
AVERAGE MAE	2.66	7.33

This summary clearly highlights that forecasts based on SASA were almost always more accurate than those realized through old-fashioned techniques. The absolute error has been below 3% in eight out of eleven predictions, and on average, the MAE is 5 points lower when using the SASA technique. The method seems to make the difference. However, to verify this finding, we provide a more systematized test of whether SASA outperforms alternative techniques through a meta-analysis of the literature on social media-based electoral forecasts.

4. META-ANALYSIS ON ELECTORAL FORECASTING: WHO BEATS WHOM?

We gathered information on several forecasts (see the Appendix for the full lists) based on web data that have been published in academic journals or that have been presented at academic conferences as well as those that are publicly available on-line and that can be easily found through a Google search using the terms "social media", "Twitter", "election" and "forecast" as queries. Overall, we collected data on 80 analyses related to elections held between 2009 and 2013 in a variety of political systems, such as France, Germany, Ireland, Italy, Japan, Netherlands, Portugal, Singapore, the United Kingdom, and the United States (see Appendix). These countries feature different rates of Internet users and different electoral systems and political institutions. Beyond parliamentary and presidential elections, we also include party primaries held to select leaders and candidates. We consider the large variety of contexts within our dataset as an added value to better investigate the strengths and limits of the different methods to monitor social media in order to forecast electoral results, as well as to assess which factors can increase (or decrease) their reliability.

The dependent variable is the previously discussed MAE of each prediction. Our main independent variable is the method adopted to forecast the election. We assess whether new or traditional sentiment analysis improves the forecasts based on the *volume of data* (our omitted category). The variable *Naïve SA* takes the value 1 when the analysis is based on old-fashioned techniques of sentiment analysis; conversely, the variable SASA takes the value 1 when the forecast is made through SASA (*i*SA or ReadMe). We then also differentiate between the two variants of SASA in a further model.

Because we are analyzing different elections across political systems, we want to control for a set of potentially confounding factors to exclude any bias. In the first model we include three main control variables. First, the variable *Internet usage* accounts for the share of Internet users across time and space, measured using World Bank data[7]. This allows one to control for the pervasiveness of Internet usage

[7] http://data.worldbank.org/indicator/IT.NET.USER.P2.

in each country. As the share of citizens that have access to the Internet grows, we would expect to observe a lower gap between predictions based on social media and actual results. If a greater share of voters has access to the web and will be active on-line, the socio-demographic traits of social media users will tend to better approximate those of the whole population of voters, thereby increasing the predictive power of the social media analysis.

In addition, we consider the role of political institutions by distinguishing the effect of different electoral systems. When elections are held in single-member plurality districts, voters may have an incentive to behave strategically: they can hide their sincere preferences and vote for a candidate that has a significant chance of becoming elected (Cox, 1997). Similarly, in plurality electoral systems (and in majority electoral systems, to a certain extent), voters may support their preferred candidate on-line while casting a strategic vote on election day. Conversely, the incentive to behave strategically is lower under proportional representation; hence, the opinions expressed on-line could be more consistent with the actual behaviour at the polls, and this can positively affect the accuracy of social media estimates. We consider this aspect through the variable *Proportional representation*, which takes the value of 1 if the election is held under proportional representation and 0 otherwise.

Furthermore, we distinguish elections in which voters have to select a monocratic position (i.e., the president, a party leader, or the mayor) rather than voting for a party list or for a potential member of the national parliament. When voters have to cast a vote on a specific candidate, the final outcome could be more easily predicted because the name of the candidate is clearly identifiable and voters' opinions are easily measurable. The variable *Personal vote* takes a value of 1 when voters have to select a single politician running for a monocratic position, and it takes a value of 0 otherwise.

Because the dependent variable MAE is a proportion whose values are bounded by 0 and 1, the assumptions required by the ordinary least-squares regression might not hold due to heteroscedasticity or because errors might not be normally distributed (Wooldridge, 2002). In addition, the predicted values might fall outside the unit interval. To address this, data have been analyzed by a fractional logit (Papke and Wooldridge, 1996). Table 3 displays the results.

Table 3: Determinants of the accuracy of social media predictions: Fractional Logit of the electoral forecast MAE (Mean Absolute Error)

	(1)	(2)	(3)
SASA	-0.998*	-1.017*	-
	(0.359)	(0.369)	
Naïve Sa	-0.070	-0.176	-0.163
	(0.252)	(0.281)	(0.285)
ReadMe	-	-	-0.863+
			(0.454)
iSA	-	-	- 1.306*
			(0.483)
Internet usage	-0.016+	-0.021*	-0.023*
	(0.009)	(0.009)	(0.009)
Proportional representation	-1.078***	-0.994**	-0.965**
	(0.303)	(0.318)	(0.317)
Personal vote	-0.470+	-0.448+	-0.436+
	(0.261)	40.247)	(0.247)
Margin of victory	-	-0.010	-0.009
		(0.008)	(0.008)
Number of candidates	-	-0.028	-0.028
		(0.046)	(0.045)
Constant	-0.812*	-0.176	-0.511
	(0.751)	(0.853)	(0.750)
N	80	80	80
Bayesian Information Criterion	-320.221	-311.542	-307.179

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The statistical analysis confirms that the method adopted to analyze the web is crucial in improving the accuracy of social-media-based electoral forecasts. All else being equal, the SASA method decreases the MAE by 5 points if compared to forecasts based on the mere number of mentions and by 4.6 points if compared to naïve SA. In contrast, *naïve SA* does not improve the accuracy of the prediction if compared to a simple volume of data.

The country's share of Internet users is relevant as well. If we increase *Internet usage* by one standard deviation from the mean, the MAE decreases by 1.4 points. This aspect could most likely be related, as previously mentioned, to the increased representativeness of social media users in countries in which Internet access is more pervasive. In addition, institutions matter. The electoral system produces a striking impact: when elections are held under proportional representation rules, the MAE decreases on average by 6.8 points. In line with our expectations, when voters have poor incentives to cast a strategic vote, the consistency between their opinion expressed on-line and their actual voting behaviour grows. Similarly, the variable *Personal vote* is negative and significant as well. Elections that ask voters to choose among candidates/persons instead of parties (or representatives of parties) tend to be more accurate by almost 3 points, most likely reflecting a "personalization effect" in electoral campaigning (for a discussion on the possible consequences of the "personalization of politics", see Kaase, 1994; Mughan, 2000).

In Model 2 of Table 3, we have performed some robustness checks by considering two additional variables. The first variable, *Margin of victory*, records the gaps between votes received by the first and second party (or coalition or candidate), thereby accounting for the degree of competitiveness of the election. We could argue that when elections are more competitive, they will attract more interest, bringing voters to "fight" on-line by expressing their vote choice. This effect, in turn, could decrease the error of the estimate. The variable *Number of candidates* records the number of parties or candidates that have been considered in the analysis to measure the MAE. Because the average error is divided by the number of parties/candidates, this value can artificially lower the MAE, and therefore we want to control for this aspect. As seen, however, neither of these two additional control variables seems to affect the accuracy of the forecast, and they do not alter the results of our analysis.

Until now, we have not explicitly distinguished between ReadMe and *i*SA forecasts, both included in the SASA category. In Model 3, we highlight this distinction by splitting the SASA category into two different dummies: ReadMe and *i*SA. As seen, both variables are negative and significant, with a magnitude, however, that is larger for *i*SA, at least in our sample. For example, compared to forecasts based on the number of mentions, on average, *i*SA reduces the MAE by more than 6 points, while ReadMe is reduced by 4.5 points.

5. CONCLUSION

Do social media analyses allow one to forecast electoral results? The present paper attempts to answer this question by adopting an SASA approach to analyze social media content. This technique (which includes the original ReadMe approach as well as the new *i*SA method) allows one to *integrate* quantitative measurements of large-n data with higher accuracy in the analysis. In a certain way, by exploiting the human coding part of the methodology, it is similar to applying focus group analyses to millions of cases. We applied SASA (both in its ReadMe and *i*SA variants) to measure the voting intention of Internet users by analyzing over 110 million comments related to elections held between 2011 and 2013 across countries and political systems, ranging from France to Italy and the United States. These estimates were then compared with actual votes and provided, on average, accurate forecasts of electoral results with a Mean Absolute Error of approximately 2.5 points.

To further explore the determinants of the accuracy of social media predictions, we compared the MAE of 80 social-media-based forecasts published in academic journals or freely available on-line. The statistical analysis reveals, first of all, that the method makes a difference. SASA increases the accuracy of the estimates by approximately 5 points when compared to forecasts based on the volume of data or on naïve SA techniques, which do not appear to be more accurate than a mere count of party/candidate mentions. Moreover, if we differentiate within the SASA category, we note that *i*SA appears to perform better than does ReadMe, at least in our sample, while ReadMe continues to outperform the remaining alternatives.

Although strongly relevant, the method is not the only factor affecting the accuracy of the prediction. First, the pervasiveness of Internet usage is crucial as well. When the rate of Internet users is higher, social-media-based predictions become more accurate. This effect is likely related to the enhanced representativeness of the web. Accordingly, we could expect that when the traits of citizens delivering comments on-line match those of the whole electorate, the web becomes a precious source of information. During the last U.S. Presidential elections, for instance, 22 million citizens declared their vote choice on-line (Pew Research, 2012). This amount of information is already relevant *per se* but can potentially become even more informative as the number of citizens active on social media grows.

The type of election also matters. The forecasts seem more accurate when citizens have to select a single candidate and cast a “vote for a person” rather than a “vote for a list”. Personal vote elections could be more straightforward and the public opinion could be easily measurable because on-line comments refer to a single well-identifiable candidate. Finally, the electoral system proves to be crucial. When the elections are held under proportional representation, social media forecasts are remarkably more precise. This effect is due to the lower incentive to cast a strategic vote. Because every vote counts in proportional electoral systems, citizens are freer to behave according to their sincere preferences. As a consequence, we observe a higher congruence between opinions expressed on-line and actual voting behaviour. Conversely, when there is an incentive to behave strategically, the analysis of the opinions expressed on-line becomes less relevant because voters may express their sincere preference on-line while casting a strategic vote at the polls.

Overall, these results seem to show that social-media-based electoral forecasts are more accurate whenever some elements strengthen the link between the opinions expressed on-line and actual behaviour. Accordingly, there are reasons to believe that, in the future, sentiment analysis may become crucial, especially if the share of people active on social media grows and if their socio-demographics traits increasingly match those of the whole population. Even in this scenario, however, the choice of an appropriate method to analyze on-line public opinion will remain decisive in producing accurate forecasts.

REFERENCES

Banks, J.S. (1991). *Signaling Games in Political Science*. Harwood Academic, Chur.

Bennett, W.L. and Segerberg, A. (2011). Digital media and the personalization of collective action: Social technology and the organization of protests against the global economic crisis. In *Information Communication and Society,* 14(6): 770–799.

Bermingham, A. and Smeaton, A. (2011). On using Twitter to monitor political sentiment and predict election results. In *Workshop on Sentiment Analysis where AI meets Psychology*, November 13, 2011, Chiang Mai, Thailand.

Ceron, A., Curini, L. and Iacus, S.M. (2013). *Social Media e Sentiment Analysis. L'evoluzione dei fenomeni sociali attraverso la Rete*. Springer, Milano

Ceron, A., Curini, L., Iacus, S.M. and Porro, G. (2014). Every tweet counts. How content analysis of social networks can improve our knowledge of citizens policy preferences. An application to Italy and France. In *New Media & Society,* 16(2): 340–358.

Ceron, A., Curini, L. and Iacus, S.M. (2015). Using sentiment analysis to monitor electoral campaigns: method matters. Evidence from the United States and Italy. In *Social Science Computer Review*, 33(1): 3–20.

Ceron, A. and d'Adda, G. (2015). E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election. In *New Media & Society,* doi: 10.1177/1461444815571915.

Choy, M., Cheong, M., Ma Nang, L. and Koo Ping, S. (2011). A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction. http://arxiv.org/abs/1108.5520. Last access: 05/10/2015.

Choy, M., Cheong, M., Ma Nang, L. and Koo Ping, S. (2012). U.S. Presidential Election 2012 Prediction using Census Corrected Twitter Model. http://arxiv.org/ftp/arxiv/papers/1211/1211.0938.pdf. Last access: 05/10/2015.

Cottle, S. (2011). Media and the Arab uprisings of 2011. In *Journalism,* 12(5): 647–659.

Cox, G. (1997). *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge University Press, New York.

De Zuniga, G.H., Puig-I-Abril, E. and Rojas, H. (2009). Weblogs, traditional sources online and political participation. In *NewMedia & Society,* 11(4):553–574.

DiGrazia, J., McKelvey, K., Bollen, J. and Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior,'*PLOS ONE,* 8(11), e79449. doi:10.1371/journal.pone.0079449

Fonseca, A. (2011). Modeling Political Opinion Dynamics Through Social Media and Multi-Agent Simulation, First Doctoral Workshop for Complexity Sciences. http://idpcc.dcti.iscte.pt/docs/Papers_1st_Doctoral_Workshop_15-6-2011/AntonioFonseca.pdf. Last access: 05/10/2015.

Franch, F. (2013). (Wisdom of the Crowds)^2: 2010 U.K. election prediction with social media. *Journal of Information Technology & Politics,* 10(1): 57–71.

Gayo-Avello, D. (2011) Don't turn social media into another "Literary Digest" poll. In *Communications of the ACM*, 54(10): 121–128.

Gayo-Avello, D. (2012). No, you cannot predict elections with Twitter. In *IEEE Internet Computing*, 16(6): 91–94

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. In *Social Science Computer Review,* 31(6): 649–679.

Giglietto, F. (2012). If Likes Were Votes: An Empirical Study on the 2011 Italian Administrative Elections. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media.*

Grimmer, J. and Stewart, B.M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. In *Political Analysis*, 21(3): 267–297.

Hopkins, D.J. and King, G. (2010). A method of automated nonparametric content analysis for social science. In *American Journal of Political Science,* 54(1): 229–247.

Jensen, M. J. and Anstead, N. (2013). Psephological investigations: Tweets, votes, and unknown unknowns in the Republican nomination process. In *Policy & Internet*, 5(2): 161–182.

Jungherr, A., Jürgens, P. and Schoen, H. (2012). Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan A, Sprenger TO, Sander PG and Welpe IM ''Predicting elections with twitter: What 140 characters reveal about political sentiment''. In *Social Science Computer Review,* 30(2): 229–234.

Kaase, M. (1994). Is there personalization in politics? Candidates and voting behavior in Germany. In *International Political Science Review*, 15(3): 211–230.

Kalampokis, E., Tambouris, E. and Tarabanis, K. (2013). Understanding the predictive power of social media. In *Internet Research,* 23(5): 544–559.

King, G. (2014). Restructuring the social sciences: Reflections from Harvard's Institute for Quantitative Social Science. In *Politics and Political Science,* 47(1): 165–172.

Larsson, A.O. and Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. In *New Media & Society,* 14(5): 729–747.

Lindsay, R. (2008). Predicting polls with Lexicon. http://www.roddylindsay.com/?p=15. Last access: 05/10/2015.

Mughan, A. (2000).'*Media and the Presidentialization of Parliamentary Elections*. Palgrave MacMillan, Basingstoke, Hants.

Metaxas, P.T., Mustafaraj, E. and Gayo-Avello, D. (2011). How (not) to predict elections". In *Proceedings of PASSAT/SocialCom 2011*, 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), IEEE Computer Society, Los Alamitos, CA, USA, 165–171.

O'Connor, B., Balasubramanyan, R., Routledge, B.R. et al. (2010). From tweets to polls: linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, 23–26 May.

Papacharissi, Z. (2002). The virtual sphere: the Internet as a public sphere. In *New Media & Society,* 4(1): 9–27.

Papke, L.E. and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401(K) plan participation rates. In *Journal of Applied Econometrics,* 11(6): 619–632.

Pew Research (2012). Social media and voting. http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_TheSocialVote_PDF.pdf. Last access: 05/10/2015.

Sanders, E. and van den Bosch, A. (2013). Relating Political Party Mentions on Twitter with Polls and Election Results. http://ceur-ws.org/Vol-986/paper_9.pdf. Last access: 05/10/2015.

Schoen, H., Gayo-Avello, D., Metaxas, P., Mustafaraj, E., Strohmaier, M. and Gloor, P. (2013). The power of prediction with social media. In *Internet Research*, 23(5): 528–543

Shi, L., Agarwal, N., Agrawal, A., Spoelstra, G. and Spolestra, J. (2012). Predicting U.S. Primary Elections with Twitter. http://snap.stanford.edu/social2012/papers/shi.pdf. Last access: 05/10/2015.

Skoric, M., Poor, N., Achananuparp, P.L., Lim, E.P. and Jiang, J. (2012). Tweets and votes: A study of the 2011 Singapore general election. In'*Proceedings of 45th Hawaii International International Conference on Systems Science* (HICSS-45 2012), IEEE Computer Society, Los Alamitos, CA, USA, 2583–2591.

Tjong Kim Sang, E. and Bos, J. (2012) Predicting the 2011 Dutch Senate election results with Twitter. Proceedings of SASN 2012, the *EACL 2012 Workshop on Semantic Analysis in Social Networks*, Avignon, France, 2012.

Tumasjan, A., Sprenger, T.O., Philipp, G.S. and Welpe, I.M. (2011). Election forecasts with Twitter. How 140 characters reflect the political landscape. In *Social Science Computer Review*, 29(4): 402–418.

Tweetminster (2010). Can Word-of-Mouth Predict the General Election Result? A Tweetminster Experiment in Predictive Modeling, http://www.scribd.com/doc/29154537/Tweetminster-Predicts. Last access: 05/10/2015.

Véronis, J. (2007). Citations dans la presse et résultats du premier tour de la présidentielle 2007. http://aixtal.blogspot.com/2007/04/2007-la-presse-fait-mieux-que-les.html. Last access: 05/10/2015.

Washington, A.L., Parra, F., Thatcher, J.B., LePrevost, K. and Morar, D. (2013). What is the correlation between Twitter, polls and the popular vote in the 2012 presidential election?, *unpublished manuscript*.

Woodly, D. (2007). New competencies in democratic communication? Blogs, agenda setting and political participation. In *Public Choice,* 134(1–2): 109–123.

Wooldridge, J.M. (2002). *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, MA.

APPENDIX

Table A1: List of forecasts considered along with the type of election, the method adopted, the Mean Absolute Error (MAE) and the source of the prediction

Election	Method	MAE	Source
German Federal Election 2009	Mentions	1.65	Tumasjan et al., 2011
German Federal Election 2009	Mentions	9.36	Jungherr et al., 2012
German Federal Election 2009	Mentions	1.57	Jungherr et al., 2012
German Federal Election 2009	Mentions	2.17	Jungherr et al., 2012
German Federal Election 2009	Mentions	4.00	Jungherr, 2013
U.S. House 2010	Mentions	21.77	Di Grazia et al., 2013
Massachusetts Senate 2010	Mentions	6.30	Metaxas et al., 2011
Colorado Senate 2010	Mentions	24.60	Metaxas et al., 2011
Nevada Senate 2010	Mentions	2.10	Metaxas et al., 2011
California Senate 2010	Mentions	3.80	Metaxas et al., 2011
Kentucky Senate 2010	Mentions	39.60	Metaxas et al., 2011
Delaware Senate 2010	Mentions	26.50	Metaxas et al., 2011
U.K. Parliament 2010	Mentions	3.00	http://www.leftfootforward.org/
Japan 2010 House	Mentions	3.38	http://senkyo.kakaricho.jp/
Japan 2010 House	Mentions	3.39	http://senkyo.kakaricho.jp/
Irland General Elections 2011	Mentions	5.58	Bermingham and Smeaton, 2011
Singapore parliamentary election 2011	Mentions	5.23	Skoric et al., 2012
Portugal Presidential 2011	Mentions	3.54	Fonseca, 2011
Portugal Presidential 2011	Mentions	10.11	Fonseca, 2011
Dutch Senate 2011	Mentions	1.33	Tjong Kim Sang and Bos, 2012
Dutch House 2012	Mentions	2.20	Sanders and van Den Bosch, 2013
Dutch House 2012	Mentions	1.90	Sanders and van Den Bosch, 2013
French Socialist Party primary 2011 (1st round)	Mentions	5.90	http://www.vanksen.fr/
French Socialist Party primary 2011 (2nd round)	Mentions	2.60	http://www.vanksen.fr/
French Presidential 2012 (1st round)	Mentions	3.38	http://www.slideshare.net/frenchweb/
French Presidential 2012 (1st round)	Mentions	6.70	http://www.vanksen.fr/
U.S. Iowa GOP primary 2012	Mentions	3.10	Jensen and Anstead, 2013
U.S. New Hampshire GOP primary 2012	Mentions	4.50	Shi et al., 2012
U.S. South Carolina GOP primary 2012	Mentions	2.76	Shi et al., 2012
U.S. Florida GOP primary 2012	Mentions	3.70	Shi et al., 2012
U.S. 2012 Popular vote	Mentions	17.90	Washington et al., 2013
Italian local election 2011	Mentions	16.05	Giglietto, 2012
Sicily Regional election 2012	Mentions	2.66	http://www.franzrusso.it/

segue

segue

Election	Method	MAE	Source
Italian Centre-Left primary 2012 (1st round)	Mentions	6.36	http://seigradi.corriere.it/
Italian Centre-Left primary 2012 (1st round)	Mentions	9.72	http://www.chefuturo.it/
Italian General election 2013	Mentions	2.81	http://vincos.it/
Italian General election 2013	Mentions	12.54	http://vincos.it/
Italian Democratic Party primary 2013	Mentions	8.51	http://www.blogmeter.it/
German Federal election 2013	Mentions	5.29	http://www.socialbakers.com/
California U.S. 2008 Presidential	Naïve SA	0.42	Gayo-Avello, 2011
Florida U.S. 2008 Presidential	Naïve SA	14.78	Gayo-Avello, 2011
Indiana U.S. 2008 Presidential	Naïve SA	14.20	Gayo-Avello, 2011
Missouri U.S. 2008 Presidential	Naïve SA	18.03	Gayo-Avello, 2011
N. Carolina U.S. 2008 Presidential	Naïve SA	16.44	Gayo-Avello, 2011
Ohio U.S. 2008 Presidential	Naïve SA	7.49	Gayo-Avello, 2011
Texas U.S. 2008 Presidential	Naïve SA	20.34	Gayo-Avello, 2011
Massachusetts Senate 2010	Naïve SA	1.20	Metaxas et al., 2011
Colorado Senate 2010	Naïve SA	12.40	Metaxas et al., 2011
Nevada Senate 2010	Naïve SA	4.70	Metaxas et al., 2011
California Senate 2010	Naïve SA	6.30	Metaxas et al., 2011
Kentucky Senate 2010	Naïve SA	1.20	Metaxas et al., 2011
Delaware Senate 2010	Naïve SA	19.80	Metaxas et al., 2011
U.K. Parliament 2010	Naïve SA	9.71	Lampos, 2012
U.K. Parliament 2010	Naïve SA	15.75	Lampos, 2012
U.K. Parliament 2010	Naïve SA	3.63	Lampos, 2012
Ireland General elections 2011	Naïve SA	3.67	Bermingham and Smeaton, 2011
Singapore 2011 Presidential	Naïve SA	6.07	Choy et al., 2011
Dutch Senate 2011	Naïve SA	2.00	Sang and Bos, 2012
London Mayoral election 2012	Naïve SA	2.47	http://www.telegraph.co.U.K./
U.S. 2012 Popular vote	Naïve SA	1.80	Washington et al., 2013
U.S. 2012 Popular vote	Naïve SA	16.00	Washington et al., 2013
U.S. 2012 Popular vote	Naïve SA	3.63	http://usatoday30.usatoday.com/
U.S. 2012 Popular vote	Naïve SA	1.29	Choy et al., 2012
U.S. 2012 Popular vote	Naïve SA	0.47	Choy et al., 2012
Italian Centre-Left primary (1st round)	Naïve SA	9.27	http://vincos.it/
Italian Centre-Left primary (1st round)	Naïve SA	8.65	http://vincos.it/
Milan mayoral election 2011 (2nd round)	SASA (ReadME)	2.60	http://voicesfromtheblogs.com/2011/10/18/palazzo-marino-ha-un-nuovo-inquilino/

segue

segue

Election	**Method**	**MAE**	**Source**
French Presidential 2012 (1st round)	SASA (ReadME)	4.65	http://voicesfromtheblogs.com/2012/04/22/presidenziali-francesi-su-twitter/
French Presidential 2012 (2nd round)	SASA (ReadME)	3.30	Ceron, Curini and Iacus, 2013
French Legislative election 2012 (1st round)	SASA (ReadME)	2.38	Ceron, Curini and Iacus, 2013
U.S. 2012 Popular vote	SASA (ReadME)	0.02	Ceron, Curini and Iacus, 2013
Italian Centre-Left primary 2012 (1st round)	SASA (*i*SA)	1.96	Ceron, Curini and Iacus, 2013
Italian Centre-Left primary 2012 (2nd round)	SASA (iSA)	1.50	Ceron, Curini and Iacus, 2013
Italian General election 2013	SASA (*i*SA)	1.62	Ceron, Curini and Iacus, 2013
Lombardy Regional election 2013	SASA (*i*SA)	1.59	http://voicesfromtheblogs.com/2013/02/25/twitter-poll-csx-avanti-di-quasi-5-punti-m5s-secondo-partito-in- lombardia-e-bagarre/
Italian Northern League primary 2013	SASA (*i*SA)	0.50	http://voicesfromtheblogs.com/2013/12/06/non-ce-solo-il-pd-anche-la-lega-ha-le-sue-primarie-e-le-vincera-salvini-almeno-su-twitter/
Italian Democratic Party primary 2013	SASA	9.17 (*i*SA)	http://voicesfromtheblogs.com/2013/12/08/primarie-pd-e-per-la-rete-the-winner-is/
Nevada Senate 2010 Senate 2010	SASA (ReadME)	1.90	http://www.fastcompany.com/
California	SASA (ReadME)	4.60	http://www.fastcompany.com/
Massachusetts Senate 2013	SASA (ReadME)	16.62	http://www.wbur.org/