

Optimising Twitter-based Political Election Prediction with Relevance and Sentiment Filters

Eric Sanders, Antal van den Bosch

CLS/CLST, Radboud University, Nijmegen

KNAW Meertens Institute, Amsterdam

e.sanders@let.ru.nl, antal.van.den.bosch@meertens.knaw.nl

Abstract

We study the relation between the number of mentions of political parties in the last weeks before the elections and the election results. In this paper we focus on the Dutch elections of the parliament in 2012 and for the provinces (and the senate) in 2011 and 2015. With raw counts, without adaptations, we achieve a mean absolute error (MAE) of 2.71% for 2011, 2.02% for 2012 and 2.89% for 2015. A set of over 17,000 tweets containing political party names were annotated by at least three annotators per tweet on ten features denoting communicative intent (including the presence of sarcasm, the message's polarity, the presence of an explicit voting endorsement or explicit voting advice, etc.). The annotations were used to create oracle (gold-standard) filters. Tweets with or without a certain majority annotation are held out from the tweet counts, with the goal of attaining lower MAEs. With a grid search we tested all combinations of filters and their responding MAE to find the best filter ensemble. It appeared that the filters show markedly different behaviour for the three elections and only a small MAE improvement is possible when optimizing on all three elections. Larger improvements for one election are possible, but result in deterioration of the MAE for the other elections.

Keywords: Large Scale Annotation, Election Prediction, Relevance Filters, Sentiment Analysis

1. Introduction

Several papers have reported on the relation of mentions of political parties in tweets and the outcome of the elections in several countries (Jungherr, 2015). There is broadly reported empirical evidence for a strong correlation between the two. This indicates that Twitter might be a good predictor of election results. However, some papers offer critical counterpoints to this claim (Gayo-Avello, 2013). Major points of critique involve the mismatch of demographics in the Twitter and election population, no or erroneous use of sentiment analysis, and ignoring the contents of the tweets. In a previous study we attempted to tackle the demographic mismatch (Sanders et al., 2016). In the present study we attempt to take into account the sentiment and contents of the tweets. To this purpose we set up a webtool for annotation of tweets in which political parties are mentioned. This resulted in a set of over 17,000 tweets classified by their communicative intent and sentiment: we asked annotators to identify sarcasm, explicit endorsements, etc. These annotations were used to create filters for exclusion of tweets with certain (combinations of) features, e.g. for removing all sarcastic tweets, or for removing all tweets in which the person posting the tweet explicitly states that he or she will not vote for a particular party.

By applying filters to our tweet counts we are in principle able to maximise the correlation between tweet mentions and election outcome. The best scoring filters can tell us something about the relevance of different kind of tweets, and the need to filter them for a proper correlation of Twitter statistics to predictions of election outcomes.

This paper is organised as follows: in Section 2 we will discuss work related to our study, in Section 3 we present the data and the prediction method used in this study. Section 4 describes the annotation process. In Section 5 we explain the filters we used and show some results. In Section 6 we draw conclusions, and we wrap up with a discussion.

2. Related Work

A strand of recent work reports on the application of sentiment analysis in the prediction of election results based on tweets. Results are mixed. Berhingam *et al.* use the recent Irish General Election as a case study for investigating the potential to model political sentiment through mining of social media (Birmingham and Smeaton, 2011). Their approach combines sentiment analysis using supervised learning and volume-based measures. They conclude that "Twitter does appear to display a predictive quality which is marginally augmented by the inclusion of sentiment analysis". Burnap *et al.* present their 'baseline' model of prediction that incorporates sentiment analysis and prior party support to generate a true forecast of parliament seat allocation of the 2015 UK General Election (Burnap *et al.*, 2016). The effect of the sentiment analysis is not clear from their analysis. Almeida *et al.* also use sentiment analysis among other techniques for their prediction of municipality elections in six Brazilian cities (Almeida *et al.*, 2015). Sentiment analysis seems to improve the results in one case, but leads to deterioration in another.

3. Material and Method

3.1. Elections

Three Dutch elections under study are the parliamentary elections of 2012 and the provincial elections of 2011 and 2015. In the latter elections, the senate is also elected, which makes it effectively a national election as well. The two types of elections, however, are different in nature. The parliamentary elections are the most important, since from the results of these the national government is formed.

3.2. Tweets

The tweets are provided by TwiNL (Tjong Kim Sang and Van den Bosch, 2013). This is a set of an estimated 40% of all Dutch tweets since December 2010. Tweets are col-

lected by querying the Twitter API with specific Dutch keywords and users that are known to be Dutch (from earlier queries). The tweets containing the name of a political party ("political tweets") are selected by pattern matching regular expressions of all political parties in the parliament; see (Sanders and Van den Bosch, 2013) for details. For our experiments, we use political tweets from ten days before, including the election day.

3.3. Prediction

The outcome of the elections are predicted by counting how often party names appear in the tweets. This is done by pattern matching of regular expressions that catch all the relevant variants of the party names. For each party the percentage of mentions in the tweets is computed and this is regarded as the prediction of the percentage in the real elections.

To determine how well the prediction is, we use the Mean Absolute Error (MAE). This is the sum of all (absolute) differences between the prediction and the election result for each party, divided by the number of parties. See equation 1 for the computation of the Mean Absolute Error.

$$MAE = \frac{\sum_{i=1}^N |Perc_{el}(i) - Perc_{tm}(i)|}{N} \quad (1)$$

where MAE is the Mean Absolute Error, $Perc_{el}(i)$, the percentage of the votes for party i in the elections, $Perc_{tm}(i)$ the percentage of mentions of party i in the tweets and N is the total number of parties.

In our baseline method we include all tweets from ten days before election day. We seek improvement in the prediction by filtering the tweets that are used and not used in the counts based on the annotations.

4. Annotation

To be able to structure and filter the raw keyword-based tweet collection we gathered, we had them annotated by crowd-sourced raters. We created a set of ten binary annotation features most of which denote a sentiment or a domain-specific communicative intent. The features and the instructions to the annotators are listed in Table 1.

These features were selected because the features or combination of features are expected to give an indication of whether the tweeter might vote for the party (s)he mentions in the tweet. E.g. a positive voting endorsement or advice is probably a strong indicator for support for a party and a sarcastic tweet is expected to show opposition of a party.

Because this annotation task is arguably subjective to some extent we aimed to have (at least) three annotations per tweet, so that we could take a majority vote for a relatively more objective annotation.

We are planning to make the data available in a standardised way. For now, the data is available by contacting the authors.

4.1. Web annotation tool

For the annotation task a web-based tool was created based on the Django framework. After logging in, one tweet is presented (as an embedded tweet requested via the Twitter API), followed by a list of annotation features that can be

ticked if appropriate. Semantically opposing features are on one line, but can both be selected as both sentiments or communicative intents can be present in a single tweet. See Figure 1 for a screen shot.

The tweets are randomly selected from a set of 705,452 tweets, 149,800 from 2011, 287,127 from 2012 and 268,526 from 2015. In order to have three annotations for the tweets, the annotation server prioritises tweets that already have two annotations (by other annotators). If these are not available, it selects tweets with one annotation and if these are not present it will select a new tweet from the complete set. The tweets from the various elections were not added to the annotation tool at the same time. In a later stage the selection procedure was changed so that the same order of tweets would be annotated for all elections.

4.2. Recruitment

Raters were recruited through the research participant system of our university. In total 17,069 tweets were annotated three times by over 500 annotators: 5375 from 2011, 6663 from 2012 and 5031 from 2015.

4.3. Annotation quality

The annotation task appeared not to be trivial. Annotators had different interpretations of the various annotation features. Some annotators were very abundant in their use of (certain) annotation values, where others were more restrained. To get some idea of the consistency of the annotations we computed inter-annotator agreement in terms of Cohen's Kappa for all annotation features for all annotator pairs. To achieve overall Kappa scores, we computed the average over all annotator pairs, and normalised for the number of tweets the pairs annotated. Table 2 lists the averages and standard deviations per feature.

For all annotation features, Kappa is 'fair' or 'moderate', except for the feature 'substantiated', for which it is 'slight'. These numbers show that there is quite some variation in the annotations, but it is far from random and taking a majority vote of the annotations is both necessary and adequate.

4.4. Analysis

Of the 17,069 tweets annotated at least three times, 2,730 contain more than one party name. For this analysis we only look at the tweets that contain only one party name. Of these remaining tweets, 4,560 belong to the 2011 elections, 5,530 to the 2012 elections and 4,249 to the 2015 elections. Table 3 shows the percentages of tweets that were annotated as 'positive' or 'negative' for each political party and the percentage of votes they got in the elections. Tweets are considered to have a certain feature if a majority selected that feature.

On average, the tweets have more often a negative sentiment than a positive one. If the positive sentiment towards a party increases from 2011 to 2012 or from 2012 to 2015, the negative sentiment decreases and vice versa in 17 of the 22 cases. Overall, the sentiment is least positive and most negative in 2015, while 2011 and 2012 have almost the same degree of sentiments. There is a remarkable difference in correlation between the two sentiments and the

Table 1: Annotation features and instructions for the annotators

Feature	Annotation Instruction
substantiated	"Check this if a statement in the tweet is supported by a (kind of) argument, The quality of the support is not important."
sarcastic	"Check this if the tweet is clearly meant to be ironic/sarcastic/cynical."
subjective	"Check this if the tweeter clearly shows his opinion."
positive	"Check this if the tweet is clearly positive. Sometimes a tweet can be positive and negative, then check both. Do not check if the tweet is neutral."
negative	"Check this if the tweet is clearly negative. Sometimes a tweet can be positive and negative, then check both. Do not check if the tweet is neutral."
positive voting endorsement	"Check this if the tweeter clearly indicates what he will vote or has voted for. Presumption what the tweeter votes is not enough. He has to state it explicitly. Sometimes a tweet can contain a positive and a negative voting endorsement, in that case, check both."
negative voting endorsement	"Check this if the tweeter clearly indicates what he will not vote or has not voted for. Presumption what the tweeter does not vote is not enough. He has to state it explicitly. Sometimes a tweet can contain a positive and a negative voting endorsement, in that case check both."
positive voting advice	"Check this if the tweeter recommends to vote for one or more specific parties. Just being positive about a party is not enough. Sometime a tweet can contain a positive and a negative voting advice, in that case check both."
negative voting advice	"Check this if the tweeter recommends not to vote for one or more specific parties. Just being negative about a party is not enough. Sometime a tweet can contain a positive and a negative voting advice, in that case check both."
no politics	"Check this if the tweet is not about politics at all (or is not in Dutch)."

Table 2: Weighted average of Cohen’s Kappa for the annotation features

Annotation feature	Average	SD
substantiated	0.15	0.21
sarcastic	0.28	0.21
subjective	0.20	0.20
positive	0.32	0.23
negative	0.40	0.21
positive voting endorsement	0.45	0.27
negative voting endorsement	0.29	0.31
positive voting advice	0.46	0.29
negative voting advice	0.31	0.33
not political	0.44	0.29
average	0.34	0.14

number of votes per party: Pearson’s R between percentage positive tweets and votes over all parties and all three elections is 0.07, while R between percentage negative tweets and votes is 0.51. The latter is puzzling. It means the more negative a party is tweeted about, the more votes it gets. For brevity reasons we only show the numbers for the ‘positive’ and ‘negative’ features. From studying all annotations we draw the following conclusions:

- Voting endorsement and advice only represent at most

10% of all tweets for most parties. Positive endorsement and advice appear more often than negative. Explicit endorsement tweets do not show the same trends as the ‘positive’ and ‘negative’ sentiment-filter did as shown above.

- Three parties are associated with a high number of non-political tweets, which could be considered errors of the keyword-based method. SP and CU are words in other languages that slipped through in our data set. This effect is smaller in 2015, possibly because of an improved language filter in TwiNL. 50PLUS can be written as 50+, which appears sometimes in a non-political context.
- The percentage of tweets annotated as ‘subjective’ is roughly between 20 and 40%. This is pretty stable over the three elections for all parties.
- The percentage of tweets annotated as ‘substantiated’ (i.e. containing some argumentation or reasoning) is always under 10% and does not show any trend across parties or elections.
- Sarcastic tweets vary between 5 and 25% for different parties. In 68% they follow the same trend as ‘negative’ over the three elections.

Figure 1: Screen capture of the webtool for annotating political tweets

POLITICAL TWEET ANNOTATOR

Geef van de volgende tweet aan welke van de onderstaande eigenschappen geldt. Kies alleen de eigenschappen waarvan het heel duidelijk is. Lees de uitleg achter de opties goed door. Als U twijfelt, kies dan niet. Probeer het zo objectief mogelijk te doen (en uw eigen politieke voorkeur uw keuze niet te laten beïnvloeden).

U heeft 300 tweets geannoteerd



<input type="checkbox"/> beargumenteerd		Kies dit als een stelling in de tweet onderbouwd wordt met een (soort van) argument. De kwaliteit van de onderbouwing is niet van belang.
<input type="checkbox"/> sarcastisch		Kies dit als de tweet duidelijk ironisch/sarcastisch/cynisch bedoeld is.
<input type="checkbox"/> subjectief		Kies dit als de tweeter duidelijk zijn eigen mening laat blijken.
<input type="checkbox"/> positief	<input type="checkbox"/> negatief	Kies dit als de tweet duidelijk positief of negatief van toon is. In een enkel geval kan het beide voorkomen, kies dan allebei. Bij een neutrale tweet niet kiezen.
<input type="checkbox"/> positieve stembekening	<input type="checkbox"/> negatieve stembekening	Kies een van de twee als de tweeter duidelijk aangeeft wat hij wel/niet zal stemmen of gestemd heeft. Vermoeden dat de tweeter op de partij stemt is niet genoeg, hij/zij moet het expliciet zeggen. In een enkel geval kan het beide voorkomen, kies dan allebei.
<input type="checkbox"/> positief stemadvies	<input type="checkbox"/> negatief stemadvies	Kies een van de twee als de tweeter aanraadt wel of niet op (een) bepaalde partij(en) te stemmen. Alleen positief of negatief over een partij spreken is niet genoeg. In een enkel geval kan het beide voorkomen, kies dan allebei.
<input type="checkbox"/> geen politiek		Kies dit als de tweet überhaupt niet over politiek gaat (of niet Nederlands is).

OK

5. Filtering

Having a substantial number of annotated tweets, we can create filters that take characteristics of tweets into account when correlating the number of party mentions with the election results. All features can have four settings in the filters:

- 1 select only tweets without this feature (NOT)
- 0 ignore this feature
- 1 select tweets with at least one of the features with this option (OR)
- 2 select only tweets with this feature (AND)

We included the tweets in which multiple parties are mentioned in our data set and use 'multiparty' as filter option to optionally exclude these tweets. All options and features can be combined.

5.1. Manual analysis

In order to perform manual tests on our data we created a web-based analysis tool with which we can set any possible filter, on the basis of which the tool computes the Mean MAE. With this tool we can observe the effect of individual features and we can try configurations of features that seem intuitive. Table 4 lists for all annotated features the effect of

filtering tweets with exclusively this feature (indicated with '+' in front of the feature name) and filtering tweets without this feature ('-'). Also a few configurations are shown as an example. Columns 2, 3 and 4 show the number of tweets that are used in the computation. Columns 5, 6 and 7 show the MAEs.

The first row with numbers represents the setting when all available data is used and the second row when all annotated data is used without filtering. The MAEs for 2011 and 2012 for all data and only annotated data are similar, but for 2015 there is a 0.25 difference. This means the MAE from the annotated data can not be taken as absolute truth, but based on the sample size we trust that trends resulting from the filters on MAEs on the annotated set will also apply to the complete set.

In two thirds of the cases, the filters with only one feature involved lead to a deterioration of the MAE. Possible explanations for this are: 1) The used filter gives a worse prediction because the collection of tweets is from a worse representation of the voters. 2) The sample size is too small for a good prediction. 3) Tweets do not reflect the way tweeters will vote at all and the effects are to be explained by something else.

The effects of the filters for 2011 and 2015 are comparable and intuitive, while those of 2012 are different and unexpected. In the latter, filters with only negative tweets and

Table 3: Percentage of tweets that are annotated as positive and negative and the election results per party for 2011, 2012 and 2015

Party	Year	'positive'	'negative'	Elections
VVD	2011	19.9	23.1	20.0
	2012	17.0	31.5	26.8
	2015	12.0	47.5	16.7
PVDA	2011	21.4	27.8	17.7
	2012	18.4	30.9	25.1
	2015	13.0	39.7	10.5
PVV	2011	5.0	48.0	12.7
	2012	9.1	36.8	10.2
	2015	8.9	42.9	12.3
SP	2011	17.3	19.9	10.4
	2012	17.6	19.5	9.7
	2015	10.6	25.9	12.2
CDA	2011	17.1	31.3	14.4
	2012	20.6	23.5	8.6
	2015	26.8	18.0	15.4
D66	2011	31.8	12.8	8.5
	2012	29.2	21.4	8.1
	2015	21.0	28.3	13.0
GL	2011	25.9	15.5	6.4
	2012	21.9	19.8	2.4
	2015	19.5	17.5	5.6
CU	2011	24.9	14.2	3.4
	2012	14.5	9.8	3.2
	2015	21.2	13.7	4.2
SGP	2011	12.8	26.9	2.2
	2012	12.2	25.0	2.1
	2015	13.9	17.6	2.9
PVDD	2011	11.3	19.4	1.9
	2012	17.2	17.2	2.0
	2015	28.1	16.2	3.6
50PLUS	2011	10.6	29.8	2.4
	2012	6.4	18.4	1.9
	2015	11.1	14.3	3.5
average	2011	17.5	28.2	
	2012	17.7	26.4	
	2015	15.7	33.6	

with leaving out positive voting endorsement or advice lead to an improvement.

Based on the effects of the individual parameters, we tested a couple of filter combinations of which the results are shown in the last three rows of the table. The first two are configurations that we expect to have a positive impact, while the latter is expected to lead to a deterioration of the results. Leaving out all sarcastic, negative and negative voting endorsement and advice tweets leads indeed to an improvement of the MAE for 2011 and 2015, but for 2012 we see the results get worse. Taking only tweets that are annotated as either positive or containing a positive voting advice or endorsement leads to an improvement for 2015 only. Leaving out all tweets that are positive or contain a positive voting advice or endorsement leads counter-intuitive to an improvement of MAE for 2012.

5.2. Automatic

In a grid search over all possible filters, we computed the lowest MAE for 2011, 2012, 2015 and the average over the three elections. A problem with filtering is that the number of remaining tweets after applying filters may be too low to compute reliable MAEs. This can be clearly seen in table 5. The table shows the best scoring filter configurations (using the number codes mentioned in the beginning of this section) for filters with a remaining number of tweets of over 0, 1000, 2000, 3000, 4000 and 5000 tweets, based on the average MAE over the three elections. Also the actual number of remaining tweets and the MAE are shown. The first column of results are those for the complete set of annotated tweets (without applying any filter). The best scoring filter is based on only 103 tweets, which is far too low for a reliable estimation. When the number of remaining tweets rises, the MAE worsens. The configuration with 1878 remaining tweets still has an MAE that is 0.31 better than the no-filter case, but the used filter makes not much sense, leaving out the tweets with most of the annotation parameters. The more tweets remain after filtering, the more sense the filter makes, but the smaller the improvement.

Because the lowest MAEs are found with complex, unrealistic filters that leave only small numbers of tweets in the data set are so unreliable, we consider filters that leave a set of at least 2,500 tweets for the three separate elections, which is about half of the annotated tweets.

For 2011 the best filter results in a MAE of 2.26 (a gain of 0.39) with 2527 remaining tweets. It selects all tweets except those with 'substantiated', 'sarcastic', 'negative', 'negative voting advice', 'positive voting endorsement' and 'negative voting endorsement'. For 2012 the best filter results in a MAE of 1.49 (a gain of 0.50) with 2529 remaining tweets. It selects all tweets including those with features 'substantiated' or 'positive' or 'negative' and without those with features 'positive voting endorsement' and 'positive voting advice'. For 2015 the best filter results in a MAE of 1.89 (a gain of 0.68) with 2587 remaining tweets. It selects all tweets except those with feature 'sarcastic', 'negative', 'positive voting advice', 'negative voting advice' and 'negative voting endorsement'.

The three filter configurations differ quite a bit from each other, which means that it is not possible to construct a filter configuration that will work for all elections and it is not possible to predict the behaviour of a filter configuration beforehand. We also notice that for all three elections the filter configurations resulting in the best results contain parameters that are counter-intuitive: it is not to be expected that leaving out all tweets that contain a positive voting advice or endorsement leads to a better prediction of the elections. This unexpected behaviour is more so for 2012 than for 2011 and 2015, a similar pattern as we saw with the individual filter features in table 4.

6. Conclusion and discussion

We created an annotated set of over 17,000 political tweets. Inter-annotator agreement showed that the annotation task is not trivial. Therefore, we took the majority vote over three annotators as ground truth. The data is available by contacting the authors. With the annotated data we created

Table 4: MAE scores for baselines, for each feature all tweets either only with this feature or without this feature and for a few configurations

Filter	#tweets			MAE		
	Year	2011	2012	2015	2011	2012
all data	149800	287127	268526	2.71	2.02	2.89
annotated data	5375	6663	5031	2.65	1.99	2.57
+substantiated	351	475	333	2.88	2.04	2.41
-substantiated	5024	6188	4698	2.65	2.01	2.63
+sarcastic	854	941	724	4.00	2.20	4.87
-sarcastic	4521	5722	4307	2.50	2.08	2.36
+subjective	1558	2169	1473	2.90	2.04	3.81
-subjective	3817	4494	3458	2.58	2.25	2.37
+positive	891	1112	756	3.36	2.87	2.48
-positive	4484	5551	4275	2.70	2.00	2.72
+negative	1519	1763	1722	4.25	1.68	4.82
-negative	3856	4900	3309	2.60	2.43	2.03
+positive voting endorsement	617	677	315	3.93	3.16	1.90
-positive voting endorsement	4758	5986	4716	2.72	1.87	2.68
+negative voting endorsement	175	175	146	5.73	3.18	2.86
-negative voting endorsement	5200	6488	4885	2.61	1.99	2.56
+positive voting advice	449	446	357	3.37	3.73	2.23
-positive voting advice	4926	6217	4674	2.70	1.91	2.66
+negative voting advice	148	176	145	5.98	2.16	3.53
-negative voting advice	5227	6487	4886	2.64	2.01	2.54
+no politics	225	442	226	5.06	8.86	2.73
-no politics	5150	6221	4805	2.73	1.89	2.68
+multiparty	820	1136	786	1.99	1.82	3.01
-multiparty	4560	5530	4249	3.23	2.28	2.54
-sarcastic -negative voting advice -negative voting endorsement -negative	3288	4265	2910	2.42	2.61	1.91
+positive OR +positive voting advice OR +positive voting endorsement	1476	1741	1106	3.38	2.86	2.10
-positive -positive voting advice -positive voting endorsement	3899	4922	3925	3.01	1.93	2.90

filters to select subsets containing tweets with features that might result in a better predictor for the election results.

Testing all possible filters and their corresponding Mean Absolute Errors showed that for 2011, 2012 and 2015, different filter configurations resulted in a lower MAE, but all containing the exclusion of features that one would not expect. The features that leads to better MAEs for 2011 and 2015 are somewhat similar and intuitive. The filter parameters for 2012 behave differently and not as expected at all. We are unable to bring MAE levels down by a substantial margin with the same filter for all elections. This may be attributed to the fact that most of our filters apply roughly equally to all political parties, rendering them ineffective for the end goal.

Manual annotated data is as close to the "truth" as we can get. Although we have a relatively big set of annotated tweets, the number of tweets that are used to compute the MAEs is rather small, especially after applying filters. In future work we would like to create a larger set of labeled data by building automatic classifiers based on the annotated data that would ideally approximate the inter-annotator agreement levels of our human annotators. On this complete set of automatically labeled tweets, we can test our filters again.

Table 5: Filter parameters used and MAE scores for best scoring configurations with a minimum number of tweets remaining after applying filters for average values over the three elections

Minimum #tweets	no filter		>0		>1000		>2000		>3000		>4000		>5000	
	#tweets	MAE	#tw	MAE	#tw	MAE	#tw	MAE	#tw	MAE	#tw	MAE	#tw	MAE
Filter	5690	2.40	103	1.80	1878	2.09	2123	2.12	3205	2.25	4206	2.29	5408	2.39
substantiated	0		1		0		0		0		0		0	
sarcastic	0		0		-1		-1		-1		-1		0	
subjective	0		-1		0		0		0		0		0	
positive	0		1		-1		-1		0		0		0	
negative	0		0		-1		-1		-1		0		0	
positive voting endorsement	0		0		-1		0		0		0		0	
negative voting endorsement	0		-1		-1		-1		-1		-1		-1	
positive voting advice	0		-1		-1		-1		0		-1		0	
negative voting advice	0		-1		-1		-1		-1		-1		-1	
no politics	0		1		-1		-1		-1		0		0	
multiparty	0		1		0		0		0		0		0	

7. Bibliographical References

- Almeida, J. M., Pappa, G. L., et al. (2015). Twitter Population Sample Bias and its Impact on Predictive Outcomes: a Case Study on Elections. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 1254–1261. ACM.
- Birmingham, A. and Smeaton, A. (2011). 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)*, pages 2–10.
- Burnap, P., Gibson, R., Sloan, L., Southern, R., and Williams, M. (2016). 140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election. *Electoral Studies*, 41:230–233.
- Gayo-Avello, D. (2013). A Meta-Analysis of State-of-the-Art Electoral Prediction from Twitter Data. *Social Science Computer Review*.
- Jungherr, A. (2015). Predictor of Electoral Success and Public Opinion at Large. In *Analyzing Political Communication with Digital Trace Data*, pages 189–210. Springer.
- Sanders, E. and Van den Bosch, A. (2013). Relating Political Party Mentions on Twitter with Polls and Election Results. In *Proceedings of DIR-2013*, pages 68–71.
- Sanders, E., de Gier, M., and van den Bosch, A. (2016). Using Demographics in Predicting Election Results with Twitter. In *International Conference on Social Informatics*, pages 259–268. Springer.
- Tjong Kim Sang, E. and Van den Bosch, A. (2013). Dealing with big data: The case of Twitter. *Computational Linguistics in the Netherlands Journal*, 3:121–134, 12/2013.