

Correlating the 2012 Dutch House of Representatives Elections based on Twitter mentions of Parties and their Party Leader

Ömer Şakar

University of Twente

P.O. Box 217, 7500AE Enschede

The Netherlands

o.f.o.sakar@student.utwente.nl

ABSTRACT

Elections are an important event within a democratic country. Different organizations try to predict the outcome of the elections based on polls. In this research, an attempt has been made to correlate the outcome of the 2012 Dutch House of Representatives elections and Twitter messages from the nomination day until the election day (44 days). The tweets have been filtered to contain only politically relevant tweets. These tweets were assigned a polarity value by a sentiment analysis library. After the tweets have been normalized in order to better depict the voters and finally an estimation of the seat distribution is made for the six largest parties. The estimation was significantly off target due to the inaccuracy of the sentiment analysis library used. A more accurate sentiment analysis library will have to be used in order to conclude if there is a correlation or not.

Keywords

election, correlation, prediction, party-based, candidate-based, Twitter, NLP, sentiment analysis, normalization

1. INTRODUCTION

During elections, there are major questions on the people's minds: which party will become the largest? How many seats will they get? Debates, leaflets, election manifestos and analysis of the current political stage are what votes will have to be based on.

In the meantime, polling organizations, such as Peil.nl and Kantar Public (formerly known as TNS Nipo), are trying to predict which party will end up with the most votes and how a coalition might be formed after. These predictions can be inaccurate to such an extent that the outcome will differ by 24 seats [1].

The inaccuracy of the polls is partially caused by the sample size of the polls. This sample size is different for the different polls and is usually a thousand to a few thousand people [3, 9]. For every poll, these people are randomly chosen from a set of correspondents in order to get a proper view of the Dutch population [9].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

27th Twente Student Conference on IT July 7th, 2017, Enschede, The Netherlands.

Copyright 2017, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

The question arises if there are other methods to predict the outcome of an election. Social media is an option to analyze since voters and politicians are both active on these platforms.

By analyzing Dutch Twitter messages which are relevant to the elections, a prediction could be made of the outcome. Messages from the nomination day (31 July 2012) to the election day (12 September 2012) of the 2012 Dutch House of Representatives are analyzed and the prediction is made for the six largest parties and for their party leader.

Since this research is *post-hoc* to the 2012 Dutch elections, it will not be a real prediction [11]. This research will also not be used to predict an election in the future. The paper will be a research to find a correlation between the outcome of the 2012 Dutch elections and Dutch Twitter messages.

The Twitter messages are filtered down from a larger set of messages. These messages are normalized in order to better represent the current voters, after which the messages are grouped based on their sentiments. And finally, a translation is made to get an overview of the number of seats the parties get.

The performance of this method is evaluated against other polling organizations and the actual outcome. The original set of tweets, which contained 208, 306, 305 Dutch tweets, is filtered down to 591, 154 tweets with 139, 341 unique users. The seat assignment has an offset of 84 relative to the final outcome of the elections (with an accuracy of 44%. Other polls have offsets between 15 and 21).

2. BACKGROUND

2.1 Twitter

Twitter is a social media platform which can be used as a microblog. Twitter users post their microblogs in the form of *tweets*. A tweet consists of at most 140 characters which could include URLs for a picture, video or other supported content. Users can follow other users, like (upvote), comment on and retweet (resend) tweets.

2.2 The Dutch elections

The Dutch House of Representatives consists of 150 seats. One seat is assigned a value based on the number of total voters. If there were 150,000 votes in total, each seat will have a value of $\frac{150,000}{150} = 1000$ (votes per seat) [4].

These seats are then distributed according to the number of votes each party gets. If each seat has a value of 1000 votes and a certain party has 1200 votes, they will get at least one seat. The additional votes (200 in the example) might get the party additional seats.

These additional seats are distributed according to the following rules. The number of votes for a party is divided by the number of seats that party already received plus 1. In our example that would be $\frac{1200}{1+1} = 600$, this we will call the *excess*. The list of parties will be sorted in descending order by this excess and the first additional seat will go to the first party on this list. If there are more additional seats to distribute, this calculation is repeated (with the newly added seat) until all seats have been filled [5].

In addition to voting for a party, a voter chooses which party member they vote for. So one might vote for the fifth candidate of "Party A". If "Party A" has only three seats, the fifth candidate could still get a seat if they have more votes than the third candidate. From previous elections, it can be observed that the party leader always has the most votes. It can be assumed that if a party gets one seat, the party leader is guaranteed to get that seat¹.

Roughly 40 days before the elections all parties have to hand in a finalized, numbered list of candidates who want to be electable. This day is called *nomination day*. Within a week the candidate lists are approved and no further adjustments can be made.

2.3 Sentiment analysis

Sentiment analysis is an area of research in Natural Language Processing (*NLP*) that determines the sentiment of a piece of text. In the context of Twitter, sentiment analysis would determine if a tweet is positive, negative or neutral with respect to a certain topic. For the classification of the tweets different classifiers, such as (Multinomial) Naive Bayes and Decision Trees can be used.

3. RELATED WORK

Table 7 has an overview of papers related to the topic of predicting elections. The table states in which country the election was, what type of election it was, how the initial data was filtered down to the relevant data, whether the data was preprocessed, whether the data was normalized and how the data is classified.

3.1 Country and type of election

There are different papers trying to predict the outcome of an election for different countries. Since elections can change from country to country, there is no way to compare them for all of these different countries. Differences between them can be the number of seats to distribute, the rules by which the winner is chosen and if a vote goes to one person or a party.

There is only one paper that deals with an election in the Netherlands [15]. The difference between the House of Representative elections (used in this paper) and the Senate elections (used in [15]) is the number of seats, which is 150 and 75 respectively.

3.2 Preprocessing

On the topic of preprocessing the tweets, most papers have a similar approach, which is removing URLs and Twitter's special character '@' and 'RT'. Some papers built upon this by removing stop words, usernames, words starting with a number, small words, punctuations, phone number and the list goes on (see Table 7, column Preprocessing).

Similarly to most papers, this one will also remove URLs and Twitter's special characters. Symbols and punctua-

¹Results from previous elections can be found on the website of the Dutch government <https://www.rijksoverheid.nl/>

tions will be removed as well, which is done in some other papers

3.3 Filtering keywords

Most papers have used either party names or candidate names to filter their data by. Some have used both and one has used candidate names in combination with events. The paper on the Dutch Senate elections only filters by party names.

This paper filters its data by the names of the six largest parties and their party leader.

3.4 Normalization

Some papers mention steps to normalize their data, other papers do not. Examples of normalization steps are grouping tweets by user, scaling the result to accommodate for some age group who are less active on Twitter and removing tweets containing multiple party names. The steps can also be data specific, for example, if the data set has to be aggregated by day instead of by the hour. And some normalization steps are mentioned to better visualize data to show in a graph.

This paper combines a few of these steps and adds one more which is explained in Section 5.3.

3.5 Classification of the tweets

Almost all papers mentioned some method to classify tweets. Sentiment analysis is used the most to determine if a tweet is positive or negative about a party/candidate. Other methods are training a Naive Bayes classifier, linear SVM (another machine learning algorithm) and using their own (mathematical) methods.

Sentiment analysis is used in this paper to determine a tweet's sentiment.

4. RESEARCH QUESTIONS

- R** Can a correlation be found between the outcome of the 2012 Dutch House of Representatives elections and Dutch tweets from the nomination day until the election day?
- R1** How can the tweets be filtered down to only contain tweets relevant to the six largest parties?
- R2** How is the sentiment of an individual tweet determined?
- R3** How can the tweets be normalized in order to better represent the Dutch voters?
- R4** How can the number of occurrences of a keyword's different sentiments be translated into votes and seats?

A flowchart of the steps taken can also be seen in Figure 1. This chart could also be seen as a data flowchart. Steps placed horizontally could be performed in any order. The substeps in R3 and R4 and the numbers in some of the steps are explained further on in Section 5 and 6.

5. RESEARCH METHOD

The initial set of tweets used in this paper has been gathered by Sang and Bos (2012) [15]. Using the Twitter streaming API, tweets with common Dutch words were filtered. This was done for every hour from January 2011 until December 2016.

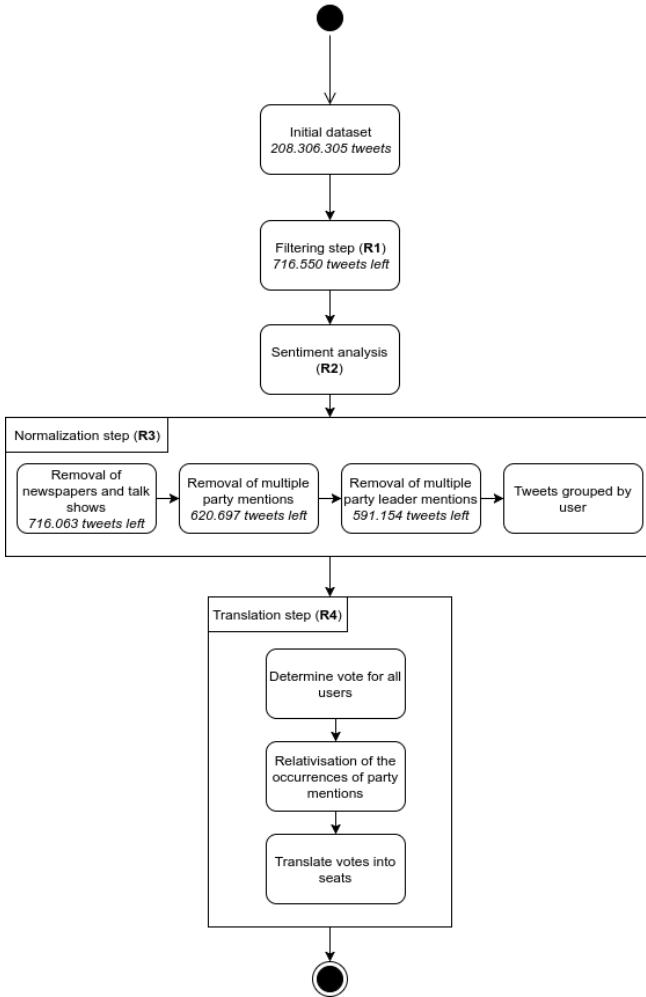


Figure 1. A flowchart of the steps taken

To narrow the number of days before the election, only the period between the nomination day (31 July 2012) and the election day (12 September 2012) is analyzed.

Whenever there is mention of selecting a random file or tweet, Python's *random* library has been used, which is a library for pseudo-random number generators.

5.1 Filtering the tweets down according to relevance (R1)

Between the nomination day and the election day, there are exactly 44 days (including the election day). Within this period, the data contains all Dutch tweets gathered. Thus an initial filtering step is needed to only work with politically relevant data. The filtering of the tweets can be summarized with these two stages: preprocessing and filtering.

5.1.1 Preprocessing

Tweets can contain URLs, special characters and symbols which make the filtering stage more difficult. URLs are defined as any string starting with 'http'. Furthermore, Twitter has two special characters which have to be removed: the user mention "@" and the hashtag "#". And finally the symbols that are removed are:

\$ " ' - , \ = + - ^ * { } [] | () : ; ? !

The reason URLs are removed is that they do not have a meaning. URLs are often shortened by Twitter's own

link shortener or by using Bitly. These shortened links are often random characters. Thus they are removed at this stage.

The two special character "@" and "#" are removed for convenience during filtering. Keeping these character makes searching for a party name more difficult since you have to deal with these possible prefixes. Since they do not add to the meaning of a tweet, this will have no effect on the outcome and are removed. The symbols are removed for the same reason as the two special characters. Note that the special characters 'RT' are not removed, since this has no effect.

As a last step of the preprocessing stage, the tweets are lowercased and multiple whitespaces are replaced by a single whitespace. This will again simplify the filtering process.

To make it clear why these special characters and symbols are removed suppose the following tweet is processed:

Ik ben die neppers in Den Haag spuugzat, wij hebben een echte leider nodig #Pechtold #D66 http://t.co/oeF58Ddi3

The result without removing the specials characters and symbols would be:

ik ben die neppers in den haag spuugzat, wij hebben een echte leider nodig #pechtold #d66

If we were looking for the keyword "pechtold", it would be necessary to keep all possible prefixes and postfixes in mind.

Besides the method of looking for a complete word, one could also check whether a keyword is a substring of a tweet. Again suppose we were looking for the keyword "SP" in the tweet above. This would give a false positive since the word "spuugzat," contains the keyword "sp" as a substring.

5.1.2 Filtering

With the preprocessed tweets, the filtering stage can start. As stated before, the list of keywords to look for is limited to the names and Twitter handles (usernames) of the six largest parties and their respective party leaders (see Figure 2). There are different ways to write some of the party names, thus all possible valid names and abbreviations are included. These keywords also go through the same preprocessing stage to remove special characters and make them lowercase.

The individual tweets are split by spaces. For each tweet, it is checked whether it contains at least one keyword and is stored if there is. The outcome of this step is a list of tweets for each hour in our chosen period. The text of these tweets are not altered (i.e. they are stored as they were before the preprocessing step).

5.2 Sentiment analysis (R2)

There are limited libraries for Dutch NLP and specifically sentiment analysis. Different libraries and services have been given a try. The service looked at is the API of text-processing.com and the libraries looked at are Frog² and BNOSAC's *pattern.nlp*³.

The API of text-processing.com was almost immediately dropped due to a daily limit of requests.

Frog is an NLP module for the Dutch language developed

²<https://languagemachines.github.io/frog/>

³<http://www.bnoscac.be/index.php/blog/54-sentiment-analysis-and-parts-of-speech-tagging-in-dutch-french-english-german-spanish-italian>

'Mark Rutte' 'Rutte' 'markrutte' 'VVD'
 'Volkspartij voor Vrijheid en Democratie'
 'Diederik Samson' 'Samson' 'diederiksamsom' 'PvdA'
 'Partij van de Arbeid'
 'Geert Wilders' 'Wilders' 'geertwilderspvv' 'PVV'
 'Partij van de vrijheid'
 'Sybrand Buma' 'Sybrand van Haersma Buma' 'Buma'
 'sybrandbuma' 'CDA' 'cdavandaag'
 'Christen Democratisch Appel'
 'EmileRoemer' 'Roemer' 'emileroemer' 'SP' 'SPnl'
 'Socialistische Partij'
 'Alexander Pechtold' 'Pechtold' 'APechtold' 'D66'
 'Democraten 66'

Figure 2. Keywords to filter by

'pauwnl' 'volkskrant' 'dwdd' 'telegraaf' 'nrc' 'adnl'
 'parool' 'metro' 'eenvandaag'

Figure 3. Usernames of the removed newspaper and talk shows

at Radboud University Nijmegen. Due to its lack of support for sentiment analysis, it was dropped.

BNOSAC's *pattern.nlp* module (written for R) is an open source library which could be used without any limits. It supports sentiment analysis for multiple languages including Dutch. Thus this package is used for sentiment analysis.

The input of the sentiment analysis step is the output of the filtering step. For each tweet, the text has been analyzed by *pattern.nlp* and has been assigned a polarity and a subjectivity. The polarity is a number between -1 and 1 and the subjectivity is a number between 0 and 1 . The tweets are again stored by the hour with their respective polarity and subjectivity.

BNOSAC does not state how the subjectivity has to be interpreted or how the polarity and the subjectivity can be combined. Thus only the polarity will be used further on. Since it is only required to know if a tweet is positive or negative, positive polarities (including 0) will be labeled 'positive' and negative polarities will be labeled 'negative'.

5.3 Normalization (R3)

Multiple steps need to be taken to normalize the data in order to get a proper view of the Dutch voters.

Firstly, the initial tweets, as well as the current tweets, are split by the hour. Since it is only necessary to know what users tweeted on daily basis, these tweets are aggregated by day.

Secondly, the current tweets contain different types of accounts, of which citizens, talk shows, and newspapers are examples. Since newspapers and talk shows cannot vote, they are removed as well. Figure 3 contains a list of the removed newspapers and talk shows.

Thirdly, in the previous section sentiment analysis was performed which produced a polarity (between -1 and 1) for each tweet. This polarity is not applicable for one word and with it, it is not possible to know to which words the polarity is based on. This could be a problem when a tweet contains multiple parties or party leader names. There is also no way to easily determine which party the tweet was positive (or negative) about. Thus tweets with multiple party names and party leader names are removed.

These two groups (parties and leaders) are removed separately in order to keep tweets which mention a party and its party leader. First tweets with multiple party names (with all possible, valid abbreviations) were removed and second, tweets with multiple party leader names (with all possible valid abbreviations) were removed. The party names and party leader names are taken from Figure 2.

And finally, since each Twitter user can only vote once, all tweets are grouped by user.

These steps are performed in this order which can also be seen in Figure 1. However, they could have been performed in any order.

5.4 Party occurrences to seats (R4)

Now that the tweets are grouped by user, a method is needed to determine the vote of one user. In order to see how the users behave in a certain timeslot, a stacked histogram is drawn (see Figure 4). 50 random users are chosen and are drawn on the X-axis. The Y-axis shows the number of positive and negative tweets for each party in the period between 6 September 2012 until 12 September 2012 (7 days). Only the positive tweet counts are of importance here, thus negative tweets are drawn black.

Additionally, the data is flattened. If a user has one positive tweet and one negative tweet about a single party, it is uncertain if the user will vote for that party. This is added up to zero tweets for that user (see Figure 4).

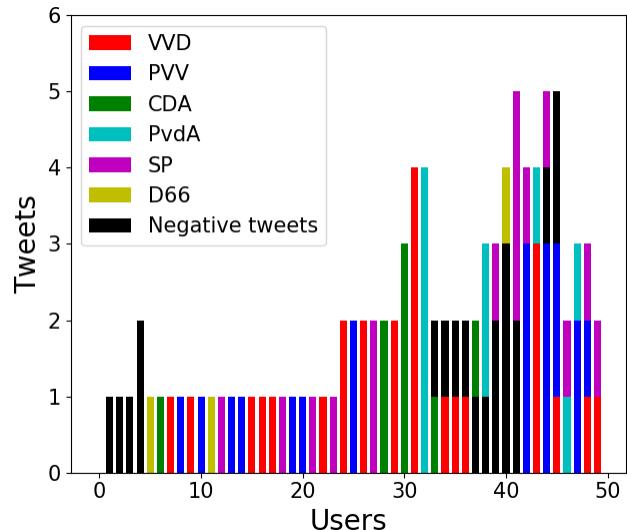


Figure 4. Positive and negative mentions of parties for 50 random users in a span of 7 days

As can be seen in Figure 4, the users can be divided into several categories.

The first few users either do not have an opinion or only have a negative opinion about some party. For these users it is not possible to determine their vote, therefore they are not included in the votes.

It can also be seen that the majority (user 5 until 41) only has a single party positively mentioned. These users will probably vote for that party.

Users 42 to 45 have multiple parties positively mentioned with one party having the most positive mentions. It is assumed that the user will vote for that party which is positively mentioned the most. The reason why this approach was taken will be explained later on.

The last few users have the same number of positive occurrences for multiple parties (i.e. a tie). For these users it is not certain which party they will vote for, therefore they are not included as votes.

This approach of taking a timeslot has one large downside. If a user has not tweeted in the chosen time span (7 days in Figure 4), their vote will not be included. Thus another approach is taken. Instead of taking tweets within a time span, the last N tweets are taken (with N still uncertain). This is done for three values of N: 5, 10 and 15.

Table 1. Number of determined votes for different N's

N	1 Party	Majority rule	Most rule	Left undecided
5	27	11	14	12/9
10	20	12	18	18/12
15	18	9	18	23/14

The last N tweets were gathered for 50 randomly chosen users. Table 1 shows for how many users it was possible to determine which party they were going to vote for. The column "1 Party" represents how many users mentioned only a single party positively, those users will vote for that party.

For the remaining users, two rules were tried. The majority rule, which states that a party has to have more than 50% of the positive mentions for a user gets the vote and the most rule, which states that the party with the most positive mentions gets the vote. These numbers can be seen in the second and third columns.

The last column describes how many users are left undecided for both the majority rule and the most rule.

It should also be noted that the number in the column left undecided for each value of N, there were 5 users who had either no opinion or no positive opinion.

It can be seen from the data in Table 1 that the majority rule is not the most effective method to determine a vote. The most rule is a more effective method and will be used if there are multiple positive party mentions.

It can also be seen from Table 1 that the difference between the different N is not large. The larger N gets, the more noise the data contains. For this reason, the result is only performed for the values $N = 5$ and $N = 10$.

5.4.1 Remaining party seats

Besides the six parties mentioned in this research, there are other parties. Out of the 150 seats to distribute, these remaining parties get a fraction as well. In order to determine how many seats the remaining parties will get, a graph is drawn with this number of seats for elections from 1981 until 2010 (see Figure 5).

The red dots represent the number of seats the remaining parties have had for a certain year. By using extrapolation an estimate can be made of the number of seats the remaining parties will get. Extrapolation needs two inputs: the current data (x- and y-values) and the order of the function which to fit against. After having tried multiple orders, the best fitting order is a simple first-order line. This line is drawn black in Figure 5. Following this line, the estimated number of seats for the remaining parties in the 2012 elections will be $15.89 \approx 16$ seats. This point is drawn green in Figure 5.

5.4.2 Translating occurrences into seats

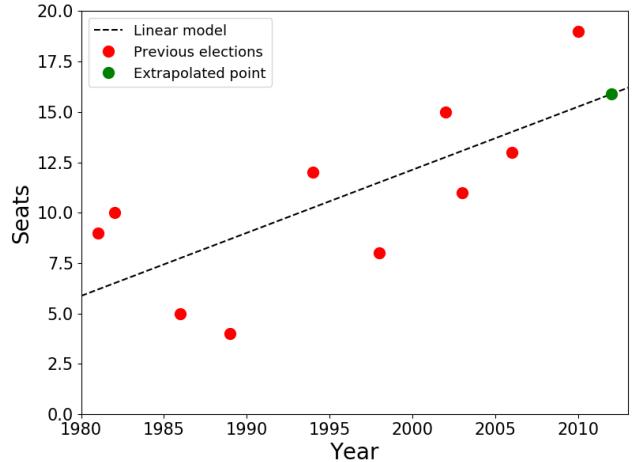


Figure 5. Seat distribution for the remaining parties

With the 'most rule' described above, the vote for one user can be determined. A translation is needed to convert the number of Twitter mentions of a party into seats. Before this translation can be performed, the input has to be relativised.

5.4.3 Relativisation

In order to fully grasp what the output looks like, suppose the following visualization (Figure 6).

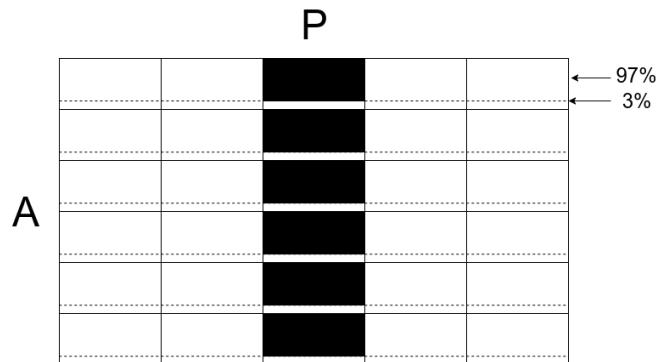


Figure 6. Visualization of the desired output

Each row represents an age group A, each column represents a party P and each cell represents the voters for party P in age group A. The sum of one column (i.e. a summation for all age groups) is the total voters for party P. However not all voters are on social media, thus the number of voters for a party in a certain age group who are also on social media is smaller.

Suppose 97% of an age group are active on social media, which means that the other 3% are not active. This shows that the summation should only contain this 97% (i.e. the number of voters on social media for party P in age group A) which is colored black.

With this in mind, let's evaluate the following equation:

$$\# \text{voters on Twitter}_P = \sum_A (\# \text{voters}_{PA} \times \% \text{voters on Twitter}_A)$$

with P = party;

A = age group;

$\# \text{voters on Twitter}_P$ = the number of P voters on Twitter (i.e. our input);

$\% \text{voters on Twitter}_A$ = the percentage of voters on Twitter in age group A;

$\# \text{voters}_{PA} = \# \text{voters}_P \times \% \text{voter}_{PA}$ = the number of P voters in age group A.

Firstly, it is a summation by age groups. Secondly, each term in the summation represents the number of voters on social media for party P in age group A, which is exactly the same as the black boxes in Figure 6. And finally, $\# \text{voters}_{PA}$ can be calculated by multiplying our unknown $\# \text{voters}_P$ with the percentage of voters in an age group out of the voters for a party (i.e. $\% \text{voter}_{PA}$).

Since the term $\# \text{voters}_P$ is multiplied for each term, the equation could also be rewritten as:

$$\# \text{voters on Twitter}_P = \# \text{voters}_P \times \sum_A (\% \text{voter}_{PA} \times \% \text{voters on Twitter}_A)$$

And since the expected output is $\# \text{voters}_P$, the equation is written as:

$$\# \text{voters}_P = \frac{\# \text{voters on Twitter}_P}{\sum_A (\% \text{voter}_{PA} \times \% \text{voters on Twitter}_A)}$$

The terms $\% \text{voter}_{PA}$ and $\% \text{voters on Twitter}_A$ can be found in two studies by CBS (Centraal Bureau voor Statistieken) [2, 6]. The relevant results can be seen in Tables 2 and 3. [2] is a study performed in 2012 and no other data could be found from previous elections, thus an estimate is done based on the number in [2] which are shown in Table 3.

Table 2. Social media usage for different age groups

Age group	Percentage on social media
18-25	97%
25-35	87%
35-45	74%
45-55	61%
55-65	50%
65-75	33%
75+	18%

The last equation will be used to relativise the input of our formula.

5.4.4 Translation of votes into seats

As stated before, there are 150 seats out of which 16 are distributed among the remaining parties, which leaves 134 seats to distribute. The total number of votes and the value of one seat is calculated. The initial number of seats are distributed by dividing the number of votes for a party by the value of a seat.

Table 3. Percentage voters in an age group out of voters for a party

	18-25	25-35	35-45	45-55	55-65	65-75	75+
VVD	1%	15%	15%	2%	2%	1%	1%
PVV	1%	15%	15%	2%	2%	1%	1%
PvdA	1%	15%	15%	2%	2%	1%	1%
SP	1%	15%	15%	2%	2%	1%	1%
D66	1%	15%	15%	2%	2%	1%	1%
CDA	0%	1%	1%	15%	15%	25%	25%

If there are seats left, they are distributed according to the following rules. The number of votes for a party is divided by the number of seats that party received initially plus 1. This is called the *excess*. The list of parties will be sorted in descending order by this excess and the first additional seat will go to the first party on this list. If there are more additional seats to distribute, this calculation is repeated (with the newly added seat) until all seats have been filled (an example is given in Section 2.2).

6. RESULTS

6.1 Filtering the tweets down according to relevance (R1)

The original data contains 208,306,305 Dutch tweets for the span of 44 days. After the filtering step, the data was filtered down to 716,550 politically relevant tweets.

In order to prove the accuracy of this method, 150 politically relevant and 150 politically irrelevant tweets were chosen randomly and labeled manually. These tweets were also labeled by the filtering algorithm and compared to the manual labeling. 42 of the 300 tweets were wrongly labeled which resulted in an accuracy of 86%.

6.2 Sentiment analysis (R2)

To prove the accuracy of BNOSAC's *pattern.nlp* module, 200 random tweets were chosen. These tweets were manually labeled positive or negative. These manual labels were compared to the polarity assigned by *pattern.nlp*. 66 of the 200 tweets were wrongly labeled which resulted in an accuracy of 67%. Upon further inspection of the false negatives and the false positives, a flaw in BNOSAC's *pattern.nlp* module was found. The module cannot correctly classify (subtle) sarcasm. This accuracy will be taken into account in the conclusion.

6.3 Normalization (R3)

The input of the normalization step was the output of the sentiment analysis step, which contains 716,550 tweets.

After the removal of the tweets from newspapers and talk shows, 716,063 tweets were left, which means 487 tweets were removed. In the end, tweets from 9 accounts were removed which would not affect the end result since they would only account for 9 votes. The list may not contain some large newspaper organization, however it will not be large enough to be significant and have an effect on the end result.

The removal of multiple party mentions resulted in 620,697 tweets and the removal of multiple party leader mentions resulted in 591,154 tweets. These two steps removed 124,909 tweets.

Grouping the tweets by user resulted in 139,341 unique users who have tweeted in the period between the nomination day and the election day.

6.4 Party occurrences to seats (R4)

For the 139,341 unique users, their votes were determined for the two values of $N = \{5, 10\}$.

Table 4. Results for N=5

	Twitter mentions	Relativised	Seats
VVD	24615	40440	31
PVV	32425	53355	40
PvdA	14911	24380	18
SP	25395	41904	32
D66	1086	1754	1
CDA	7438	16372	12
Ties	15954	-	-
Negatives	17516	-	-
No parties	1	-	-

Table 5. Results for N=10

	Twitter mentions	Relativised	Seats
VVD	24729	40440	31
PVV	32626	53355	40
PvdA	14908	24380	18
SP	25624	41904	32
D66	1072	1754	1
CDA	7449	16372	12
Ties	15490	-	-
Negatives	17442	-	-
No parties	1	-	-

Even though the number of Twitter mentions for each party is slightly different for both values of N , the seats each party receives is exactly the same.

Table 6 shows the polls on 11 September 2012 for six different polling organizations. These polls are the final polls before the elections. The six polling organizations are Peilingwijzer (PW), De Stemming (DS), Maurice de Hond (MdH), Politieke Barometer (PB) and TNS NIPO (TN). The Twitter column is the outcome of this research and the Elections column is the actual outcome of the 2012 elections. The last row shows the offset (absolute difference) between the polls and the election.

Table 6. Comparison with other polls and their offsets

	PW	DS	MdH	PB	TN	Twitter	Elections
VVD	35-37	35	36	37	35	31	41
PVV	17-19	17	18	17	17	40	15
PvdA	35-37	34	36	36	34	18	38
SP	19-21	22	20	21	21	32	15
D66	10-12	11	11	10	13	1	12
CDA	11-13	12	12	13	12	12	13
Offset	15-19	21	17	16	20	84	0

7. CONCLUSION

The offset of our research poll is 84 seats which means that only 66 seats were assigned correctly (see Table 6). This offset is approximately four times larger than the other polls, which is significant. This results in an accuracy of $\frac{66}{150} = 0.44 \cong 44\%$.

However, this is not necessarily a bad result. Keeping in mind the accuracy our filtering step (86%) and the sen-

timent analysis library (67%) it is expected to be inaccurate to some degree. Suppose the method would have been absolutely perfect (i.e. an accuracy of 100%) with the drawbacks being the filtering step and the sentiment analysis library, it is expected to have an accuracy of $86\% \times 67\% = 57.6\%$. If the current 44% accuracy is compared against this 57.6% expected accuracy, the rest of the method seems valid while still needing improvement.

In conclusion, the answer to the question if there is a correlation between Dutch tweets and the outcome of the 2012 Dutch House of Representatives elections seems to be left unanswered. The experiment will have to be repeated with a more accurate sentiment analysis library in order to state with certainty if there is a correlation or not.

8. DISCUSSION

8.1 Sentiment analysis

To address the elephant in the room, the sentiment analysis has to be improved. With only an accuracy of 66%, it is not reliable enough. Upon inspection of the false positives and false negatives, it seems like this low accuracy is largely caused by sarcasm in the tweets. There are multiple ways to solve this problem.

- Choosing a sentiment analysis library that can deal with sarcasm (which is the most straightforward method).
- Combining the current sentiment analysis library with a sarcasm detector.
- Combining the two outputs (polarity and subjectivity) of the sentiment analysis library. Perhaps there is a correlation between these and sarcasm.
- Correct the number of positives by estimating how many false positives there are and removing them.
- Instead of using sentiment analysis, an estimation of the percentage of positive tweets can be made for each party.
- Using machine learning algorithms such as Naive Bayes to make your own classifier.

The points mentioned below will assume that the sentiment analysis step works accurately.

8.2 Determining a vote

At the moment the method to determine a vote is simple. If a user mentions only a single party positively, their vote will go to that party. If there are multiple parties mentioned with one party that has the most positive mentions, that party will get the vote. In all other situation, a vote cannot be determined. These situations will be discussed separately below.

The easiest two situations are when there is no vote (positive and negative mentions add up to zero) and if there are only negative votes. Currently, by only looking at positive mentions, it is not possible to determine these votes.

The most important situation is when there are multiple parties with equal positive mentions (i.e. ties). These ties make up a significant piece of the votes as can be seen in Tables 4 and 5. Being able to determine these votes, could improve the methodology.

An entirely different way to determine votes is in fractions of votes. If party A is mentioned 5 times positively and

party B is mentioned 4 times positively, party A would get $\frac{5}{9}$ of a vote and party B would get $\frac{4}{9}$ of a vote. This method has to be evaluated as well.

8.3 Estimating data

Tables 2 and 3 have data which are estimations based on two studies by CBS. These studies are from 2012 which means that this data was not available at the time of the elections. A better way has to be found to estimate the data in these tables.

8.4 Optimum N

In Section 5.4 it is stated how N was determined. This N stands for the last N tweets a user has tweeted. However, this is not the correct method to determine N. The methodology in Section 5.4 will have to be performed for more values of N which after the optimum value for N can be chosen.

8.5 Actual prediction

Due to the time constraint, the methodology is not applied on other elections. In order to prove that this methodology works, it still needs to be tested on another election. This can be an election from the past (the 2017 elections for example) which will be another correlation or an election to come which will be a prediction.

8.6 Comparison with related work

Compared to results from the papers in the related work, the results in this paper are significantly worse. In order to do this paper justice, the methodology will have to be improved with the steps mentioned above. Only then will the comparison be fair.

9. REFERENCES

- [1] Exitpoll Slotpeilingen en exitpoll vergeleken met verkiezingsuitslag 12 september 2012.
<http://www.peilloos.nl/peilingen-sinds-val-kabinet-2012/recent/>. Accessed: 03-05-2017.
- [2] Het profiel van het electoraat in 2012.
<https://www.cbs.nl/nl-nl/achtergrond/2017/08/het-profiel-van-het-electoraat-in-2012>. Accessed: 21-06-2017.
- [3] Hoe komt De Stemming tot stand?
<http://stemming.eenvandaag.nl/?peilingId=79>. Accessed: 27-06-2017.
- [4] Kiesdrempel, kiesdeler en voorkeurdrempel.
<https://www.kiesraad.nl/verkiezingen/inhoud/tweede-kamer/uitslagen/kiesdrempel-kiesdeler-en-voordeurdrempel>. Accessed: 09-05-2017.
- [5] Zetelverdeling over partijen. <https://www.kiesraad.nl/verkiezingen/inhoud/tweede-kamer/uitslagen/zetelverdeling-over-partijen>. Accessed: 10-05-2017.
- [6] Zeven op de tien gebruiken sociale media.
<https://www.cbs.nl/nl-nl/nieuws/2013/40/zeven-op-de-tien-gebruiken-sociale-media>. Accessed: 09-05-2017.
- [7] A. Bermingham and A. F. Smeaton. On using Twitter to monitor political sentiment and predict election results. 2011.
- [8] A. Ceron, L. Curini, S. M. Iacus, and G. Porro. 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 & Society*, 16(2):340–358, 2014.
- [9] M. de Hond. Onderzoeksopzet.
<https://home.noties.nl/peil/onderzoeksopzet/>. Accessed: 10-05-2017.
- [10] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLoS one*, 8(11):e79449, 2013.
- [11] D. Gayo-Avello. "I wanted to predict elections with Twitter and all I got was this lousy paper"—a balanced survey on election prediction using Twitter data. *arXiv preprint arXiv:1204.6441*, 2012.
- [12] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani. Buzzer detection and sentiment analysis for predicting presidential election results in a Twitter Nation. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*, pages 1348–1353. IEEE, 2015.
- [13] R. Jose and V. S. Chooralil. Prediction of election result by enhanced sentiment analysis on Twitter data using word sense disambiguation. In *Control Communication & Computing India (ICCC), 2015 International Conference on*, pages 638–641. IEEE, 2015.
- [14] M. A. Razzaq, A. M. Qamar, and H. S. M. Bilal. Prediction and analysis of Pakistan election 2013 based on sentiment analysis. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 700–703. IEEE, 2014.
- [15] E. T. K. Sang and J. Bos. Predicting the 2011 Dutch senate election results with Twitter. In *Proceedings of the workshop on semantic analysis in social media*, pages 53–60. Association for Computational Linguistics, 2012.
- [16] J. Smailović, J. Kranjc, M. Grčar, M. Žnidarsić, and I. Mozetič. Monitoring the Twitter sentiment during the Bulgarian elections. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–10. IEEE, 2015.
- [17] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10(1):178–185, 2010.
- [18] E. Tunggawan and Y. E. Soelistio. And the winner is . . . : Bayesian Twitter-based prediction on 2016 U.S. presidential election. In *2016 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pages 33–37, Oct 2016.
- [19] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, pages 115–120. Association for Computational Linguistics, 2012.
- [20] Z. Xie, G. Liu, J. Wu, L. Wang, and C. Liu. Wisdom of fusion: Prediction of 2016 Taiwan election with heterogeneous big data. In *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–6, June 2016.

Table 7. An overview of the related work

Study	Country	Type of election	Preprocessing	Filtered by	Normalization	Classification
[16]	Bulgaria	Parliamentary	Tokenization, removal of stop words, lemmatization and n-gram construction	Party names and candidates	✓	Linear SVM
[13]	Delhi (India)	Legislative	Removal of URLs, private usernames and hashtags	Candidates' name	✗	SentiWordNet (SWN) and WordNet
[8]	France	Presidential and legislative	-	Party names and candidates	✗	Supervised sentiment analysis
[17]	Germany	Parliament	-	Party names	✓	Sentiment analysis
[12]	Indonesia	Presidential	Removal of the '@' character, URLs and retweet characters 'RT'	Candidate names	✓	Sentiment Lexicon
[7]	Irish	General elections	Removal of topic terms, usernames and URLs	Party names	✗	Adaboost MNB
[15]	Netherlands	Senate	-	Party names	✓	Sentiment analysis
[14]	Pakistan	General elections	Removal of duplicate tweets, URLs, extra whitespace, repeated words, words starting with a number, small words and punctuations	Party names and candidates	✓	Different classifiers such as Naive Bayes, k-nearest neighbors, and Prind
[20]	Taiwan	General elections	-	Party names	✗	Sentiment analysis and Naive Bayes
[19]	USA	Presidential	Tokenizing with Christopher Potts' Twitter tokenizer which can handle URLs, emoticons, phone numbers, HTML tags, mentions, hashtags, repetition of symbols, Unicode characters, fractions and decimals	Candidate names and events	✓	Naive Bayes
[10]	USA	House of Representatives	-	Candidate names	✓	-
[18]	USA	Presidential	Removal of URLs, pictures	Candidates' name	✗	Naive Bayes