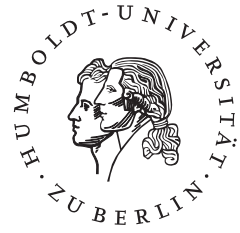


HUMBOLDT-UNIVERSITÄT ZU BERLIN



MATHEMATISCH NATURWISSENSCHAFTLICHE FAKULTÄT II
INSTITUT FÜR INFORMATIK

Lässt sich der Ausgang von politischen Wahlen durch die Analyse von Twitterdaten vorhersagen?

Exposé zur Diplomarbeit

Betreuer: ULF LESER

Martin Beckmann

17. Januar 2014

Inhaltsverzeichnis

1	Motivation	2
2	Nutzung von Twitter durch Politiker	3
3	Twitter als Prognosewerkzeug	4
3.1	Verwandte Arbeiten	4
3.2	Schritt 1: Wahl der Daten	6
3.3	Schritt 2: Verarbeitung der Daten	7
3.4	Schritt 3: Post-Processing	7
3.5	Anwendung und Evaluation	8



Barack Obama

@BarackObama

Four more years. pic.twitter.com/bAJE6Vom

Tweet 1: Nachricht von Barack zu seiner Wiederwahl im November 2012

1 Motivation

Aus unserer heutigen Zeit ist das Internet nicht mehr wegzudenken. Viele Menschen sind nicht nur zu hause online, sondern auch unterwegs durch Smartphones. Vernetzt sind die Menschen dabei durch soziale Netzwerke. Sie bieten die Möglichkeit sich selbst darzustellen und mit Leuten in Kontakt zu treten. Die beiden größten Vertreter von Social Networks Platforms sind Facebook und Twitter. Zu ihren Nutzern gehören dabei nicht nur Privatpersonen, selbst Firmen verfügen meist über ein Profil bei einem der Anbieter. Dass dieses Mittel das öffentliche Ansehen steigern kann, ist auch schon in der Politik angekommen. Barack Obama nutzte schon bei seiner Präsidentschaftskandidatur 2008 soziale Netzwerke, um mit potenziellen Wählern in Kontakt zu treten. Auch 2012 forderte er seine Anhänger über diese Kanäle auf, wählen zu gehen. Nach seiner erfolgreichen Wiederwahl veröffentlichte er am 6. November 2012 ein Foto mit dem Titel „Four more years“, dass ihn zusammen mit seiner Frau zeigte (siehe Tweet 1). Dieses Foto wurde innerhalb von wenigen Stunden auf Twitter zum bekanntesten Bild aller Zeiten mit den meisten Retweets[Jup12].

2 Nutzung von Twitter durch Politiker

Auch in der deutschen Politik haben Twitter und Facebook schon Wurzeln geschlagen. Die Vertreter der einzelnen Parteien nutzten Twitter, um vor der Bundestagswahl 2013 Meinungen zu äußern und Wähler anzusprechen. So twitterten unter den Namen @peersteinbrueck, @GregorGysi, @SWagenknecht, @JTrittin und @GoeringEckardt die Spitzenkandidaten der SPD, Die Linke (insgesamt hatte die Linke acht Spitzenkandidaten, hier wurden nur zwei gelistet) und Bündnis 90/Die Grünen. Interessant ist, dass gerade die CDU-Spitzenkandidatin und Bundeskanzlerin Angela Merkel und auch der Spitzenkandidat der FDP, Rainer Brüderle, über keinen aktiven Twitteraccount verfügt. Dafür existieren aber, wie bei den anderen Parteien auch, gesamtparteiliche Twitteraccounts mit den Namen @cdu und @fdp.

Wie es ansonsten mit der Aktivität der deutschen Politiker auf Twitter aussieht, werde ich im ersten Teil meiner Diplomarbeit betrachten. Dabei geht es vor allem um die Fragen:

- Welche Politiker besitzen Accounts bei Twitter?
- Wie aktiv werden diese genutzt?
- Gibt es Unterschiede bei den einzelnen Parteien?
- Gibt es eine Beziehung zwischen der Aktivität bei Twitter und dem Alter der Politiker oder der Stellung innerhalb der Partei?
- Können die Tweets der Politiker dazu genutzt werden um Anhänger der einzelnen Parteien zu identifizieren und somit ein Stimmungsbild auf Twitter zu erstellen?
- Kann ein solches Stimmungsbild dazu genutzt werden um politische Wahlergebnisse vorherzusagen?



Sebastian Finsel

@finsel

Die Europawahlen haben begonnen! Meine Empfehlung: Am Sonntag: das soziale Europa wählen! #spd+

Tweet 2: Ein Tweet aus dem Jahr 2009, in dem zum SPD-Wählen aufgefordert wird.

3 Twitter als Prognosewerkzeug

Der zweite Teil meiner Diplomarbeit beschäftigt sich mit einem weitergehenden Thema, der Nutzung von Twitter als Prognosewerkzeug. Es geht darum, die Tweets, die von Twitter frei zugänglich gemacht werden zu nutzen, um reale Ereignisse vorherzusagen oder zu rekonstruieren. Verschiedene Veröffentlichungen (Siehe [PD11], [BMZ11], [SOM10]) haben gezeigt, dass dies in unterschiedlichen Bereichen durchaus praktikabel ist und teilweise verblüffend ähnliche Ergebnisse liefert wie wesentlich aufwändigere Verfahren.

3.1 Verwandte Arbeiten

Es stellt sich nun die Frage, ob solche Verfahren auch auf politische Meinungen anwendbar sind. Lassen sich Tweets automatisch auf Informationen untersuchen, um so ein politisches Stimmungsbild zu erstellen? Und stimmt dieses Stimmungsbild auch mit dem überein, dass in der realen Welt existiert?

Mit diesen Fragestellungen haben sich schon verschiedene Autoren beschäftigt. Dabei kamen sie zu sehr unterschiedlichen Ergebnissen. So wurde in [TSSW10] gezeigt, dass eine Vorhersage von Wahlergebnissen anhand von Twitterdaten möglich ist und die Ergebnisse mit anderen Prognosen mithalten können. Dazu wurden die Nennungen der einzelnen Parteien (siehe Tweet 2) in einem gewissen Zeitraum vor der Bundestagswahl 2009 in Deutschland in Tweets gezählt und daraus eine Stimmenverteilung ermittelt. Beim Vergleich mit dem Ausgang der eigentlichen Wahl ergab sich ein mean absolute error (MAE) von 1,65 %, was nah an den Prognosen der Forschungsgruppe Wahlen (MAE: 1,48 %) und Infratest/dimap (1,28 %) liegt.

Als direkte Antwort darauf veröffentlichten A. Jungherr et al. ein Jahr später einen Artikel [JJS12]. Darin legten sie nahe, dass die veröffentlichten Ergebnisse nur durch gezielte Wahl bestimmter Parameter möglich war. Wäre beispielsweise ein anderer Zeitraum gewählt worden, so wären die Resultate möglicherweise schlechter ausgefallen (MAE von bis zu 3,34 %). Ein weitaus wichtiger Punkt, der gemacht wurde, war allerdings, dass eine Berücksichtigung der neu angetretenen Piratenpartei die Ergebnisse drastisch geändert hätten. Diese wurden in dem Originalpaper nicht einbezogen, hätten aber nach dem vorgestellten Verfahren ein Ergebnis von 34,8 % der Stimmen auf Twitter bekommen.

Dies steht in keinem Verhältnis zu den 2,1 %, die die Piratenpartei bei der eigentlichen Bundestagswahl 2009 erhielt. Auch wenn ein Teil dieser Stimmen womöglich auf die Ambiguität des Wortes Piraten zurückzuführen ist, lässt sich der hohe Anteil bei Twitter darauf zurückführen, dass sich die Piratenpartei mit seinem Programm vor allem an junge, internet-affine Menschen richtet.

Eine wichtige Besonderheit, die bei der Vorhersagen von realen Gegebenheiten anhand von sozialen Netzwerken nicht vergessen werden darf, ist, dass die Nutzer solcher Netzwerke nur einen Teil der Bevölkerung widerspiegeln. Dass dieser Teil keine zufällige Auswahl der Bevölkerung ist, wurde in [MLA⁺11] gezeigt. Darin wurden die Nutzer von Twitter nach verschiedenen Merkmalen in Gruppen eingeteilt. Es stellte sich heraus, dass Männer überrepräsentiert sind sowie Menschen, die in Großstädten wohnen. In [NGTM13] konnte am Beispiel von niederländischen Twitterusern gezeigt werden, dass vor allem auch das Nutzeralter eine starke statistische Verzerrung aufweist. So waren rund 60 % der untersuchten Profile von Nutzern unter 20 Jahren, 26 % von 20 bis 40 Jährigen und nur knapp 14 % der Nutzer waren über 40 Jahre alt. Die logische Schlussfolgerung, die auch von [MLA⁺11] gezogen wird, ist, dass ein Ansatz, der versucht Vorhersagen anhand von Twitter zu machen, die besondere Demographie sozialer Netzwerke berücksichtigen muss und somit post-hoc Korrekturen an den Resultaten vornehmen sollte.

Geht man davon aus, dass die Altersverteilung deutscher Twitteruser ähnlich zu der oben genannten ist und vergleicht diese dann mit der Stimmverteilung bei den repräsentativen Wahlergebnissen des Bundeswahlleiters von 2009 ¹ (Ergebnisse zur Bundestagswahl 2013 erscheinen erst am 27. Januar 2014), so zeigt sich eine starke Divergenz. Einerseits dürfen in Deutschland bei der Bundestagswahl nur volljährige Bürger wählen (ab 18 Jahren), so dass ein Großteil der bei Twitter stark repräsentierten Gruppe der unter 20 Jährigen überhaupt kein Stimmrecht hat. Eine Hypothese, die es in diesem Zusammenhang zu klären gilt, ist, ob Twitteruser, die selbst noch nicht wahlberechtigt sind, auch keine Tweets über die Parteien veröffentlichen. Andererseits bekommen die Parteien über 60 % ihrer Stimmen von Wählern über 45 Jahren, die aber bei Twitter nur mit 14 % vertreten sind. Es gibt noch einige weitere Publikationen zum Thema Nutzung von Twitter als Prognosewerkzeug für politische Wahlen, die in der Arbeit dann genauer betrachtet werden, wie beispielsweise [SB12], [MMGA11] und [Jun13b]. Einen guten Überblick zu dem Thema bietet auch [GA12].

Auch in den Medien wurde das Thema speziell vor der Bundestagswahl 2013 vermehrt erörtert. Es gibt beispielsweise Artikel in „Die Welt“ [BG13], „Spiegel Online“ [Jun13a] und „Zeit Online“ [dpa13]. Dies zeigt, dass auch die Öffentlichkeit Interesse an diesem Thema hat. Mit dem Wahl-O-Meter ² und dem Twitterbarometer ³ existieren im Internet

¹Siehe http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_09/veroeffentlichungen/repraesentative/index.html

²Siehe <http://wahl-o-meter.com/Willkommen.html>

³Siehe <http://twitterbarometer.de/>

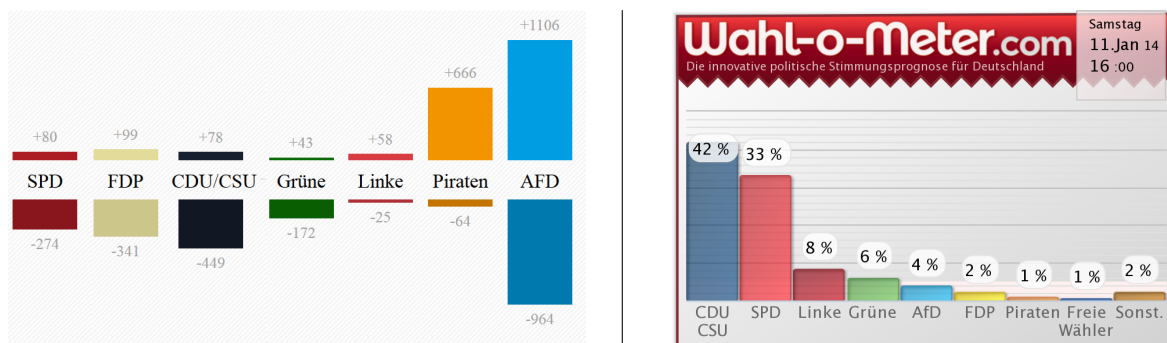


Abbildung 1: Eine Gegenüberstellung der Resultate des Twitterbarometers (links - <http://twitterbarometer.de/widget/bar500x350> von 11. Januar 2013) und des Wahl-O-Meters (rechts - http://wahl-o-meter.com/exportImg/wom_party_pct_current.svg vom 11. Januar 2013). Während in der linken Darstellung die Piratenpartei und die AFD stark überrepräsentiert sind, wurde dieser Fehler beim Wahl-O-Meter durch Post-Processing-Schritte behoben.

zwei Seiten, die in Echtzeit Tweets untersuchen und daraus ein politisches Stimmungsbild erzeugen. Das Twitterbarometer zählt dabei lediglich die Vorkommen von positiven und negativen Hashtags der Parteienamen (beispielsweise #CDU- und #SPD+) und gibt die Zahlen ohne weitere Nachbearbeitung an. Im Gegensatz dazu wird beim Wahl-O-Meter ein Post-Processing durchgeführt, mit dem die Daten geglättet werden und das Onlineverhalten unterschiedlicher Wählergruppen berücksichtigt wird wie beispielsweise der „Netzpräsenz und Social-Web-Aktivität von Piraten vs. CDU/CSU-Wählern“. Details zu dem durchgeführten Post-Processing wurden allerdings nicht genannt. Ein Vergleich der Resultate beider Seiten zeigt erneut, dass eine Nutzung der Twitterdaten ohne weitere Nachbearbeitung schlechtere Ergebnisse liefert (siehe Abbildung 1).

Die Frage, die sich nun stellt, ist, wie ein sinnvolles Verfahren zur Annäherung des politischen Meinungsbildes im Internet und dem in der realen Welt aussehen kann. Ich werde diese Frage im zweiten Teil meiner Arbeit bearbeiten und einen Ansatz dazu liefern. Ein solcher Ansatz lässt sich grundsätzlich in drei Schritte einteilen.

3.2 Schritt 1: Wahl der Daten

Wichtig ist, welches Verfahren gewählt wird, um die Daten zu sammeln, die genutzt werden sollen. Das Verfahren kann dabei einfach sein, wie beispielsweise die Nutzung aller Tweets mit den Hashtags der Parteien ([TSSW10], [JJS12]). Mein Ansatz wird hingegen komplexer sein und auf Ergebnisse aus dem ersten Teil (Siehe Kapitel 2) zurückgreifen. Darin wird das Verhalten von Politikern auf Twitter analysiert. Nutzt man die Tweets dieser Politiker, um einen Classifier zu trainieren, so lassen sich Tweets von anderen Nutzern danach kategorisieren, welchen Parteien sie ähnlich sind. Ein Tweet t wird dabei einer Partei c zugeordnet, bei der die bedingte Wahrscheinlichkeit $p(c|t)$ am größten ist unter allen Parteien c_i : Nützlich bei der Wahl eines Naive-Bayes-Klassifikators für das Verfahren

ist, dass dabei leicht Wörter gefunden werden können, die besonders charakteristisch für bestimmte Parteien sind (ein hoher Wert für $p(w|c)$). Diese können dann ähnlich zu den Hashtags der Parteien dazu genutzt werden, weitere Tweets zu finden, die solche Wörter beinhalten.

3.3 Schritt 2: Verarbeitung der Daten

Der zweite Schritt für die Erstellung von Vorhersagen durch Twitterdaten besteht in der Überführung der Sammlung an Tweets in numerische Werte. In [TSSW10] und [JJS12] wurde dies beispielsweise durch die Zählung der Hashtags der Parteien durchgeführt. In anderen Fällen [SB12] wurde für jeden Nutzer nur einer der Tweets verwendet, statt alle Tweets zu benutzen. Dadurch sollte eine Annäherung an die Beschaffenheiten in der Realität erreicht werden, in der Wähler bei einer Wahl auch nur eine Stimme besitzen. Somit soll Nutzern, die besonders viele Tweets schreiben, der gleiche Einfluss gegeben werden, wie Nutzern, die innerhalb des Zeitraums nur einen Tweet verfasst haben. Ein Ansatz, der bisher noch nicht verfolgt wurde, besteht darin, jedem Nutzer nur eine Stimme zu geben, diese aber nicht anhand seines ersten Tweets zu vergeben, sondern auf die Parteien zu verteilen, denen seine Tweets zugeordnet wurden. So wird einerseits weiterhin gewährleistet, dass alle Nutzer den gleichen Einfluss haben. Andererseits können Nutzer aber mehrere Parteien unterstützen, falls sie beispielsweise unschlüssig sind, welche der Parteien sie wählen würden.

3.4 Schritt 3: Post-Processing

Der letzte wichtige Schritt des Verfahrens besteht in der Nachbearbeitung der errechneten Werte. Dadurch können Fehler, die in den ersten beiden Schritten entstanden sein könnten, behoben werden. In erster Linie soll dadurch eine Verzerrung der Daten reduziert werden, die durch die Wahl der verwendeten Tweets und insbesondere durch die Wahl von Twitter als Datenquelle kommt. Wie schon im Abschnitt 3.1 geschildert, besteht eine starke Divergenz zwischen der Twitter-Nutzerschaft und den eigentlich gewünschten Wählern in Deutschland. Diese Fehlstellung in der Stichprobe kann durch gezielte Anpassung der Ergebnisse umgangen werden. Dazu sollen die Tweets nach ihrem Erstellungsdatum in Gruppen eingeteilt werden mit Abständen von circa zwei Wochen. In einem solchen Abstand werden seit 1991 durch die Forschungsgruppe Wahlen Umfrageergebnisse veröffentlicht zu der Fragestellung „Wenn am nächsten Sonntag wirklich Bundestagswahl wäre ...“ (Siehe Abbildung 2). Geht man davon aus, dass die Stichprobe bei diesen Umfragen näher am eigentlichen Bevölkerungsschnitt liegt, so bietet es sich an, die Ergebnisse durch Twitter an diese Daten anzunähern. Zu überprüfen ist, ob es einen Zusammenhang zwischen den Ergebnissen von Twitter und der Forschungsgruppe Wahlen für die einzelnen Parteien gibt. Ist eine solche Korrelation gefunden, so kann diese genutzt werden, um

Projektion seit 01/1991:

Wenn am nächsten Sonntag wirklich Bundestagswahl wäre ...

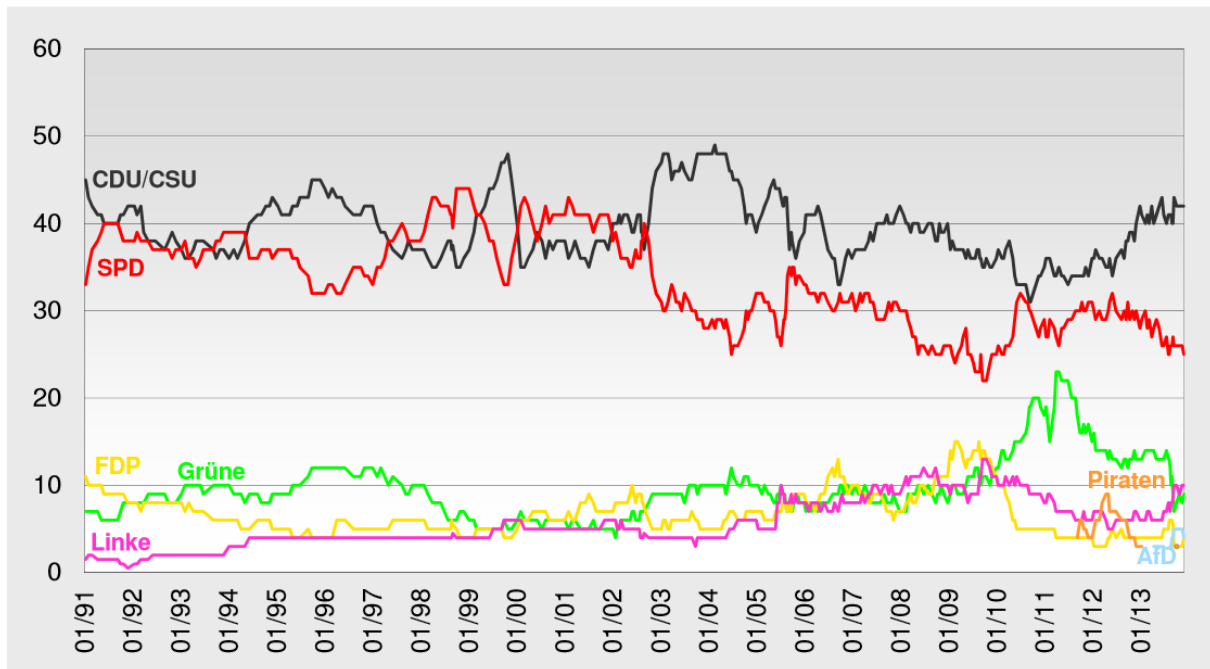


Abbildung 2: Graphische Darstellung der Umfrageergebnisse der Forschungsgruppe Wahlen zum Thema „Wenn am nächsten Sonntag wirklich Bundestagswahl wäre ...“ (Siehe http://www.forschungsgruppe.de/Umfragen/Politbarometer/Langzeitentwicklung_-_Themen_im_Ueberblick/Politik_I/)

die Ergebnisse durch Twitter anzupassen, so dass sie den realen Gegebenheiten näher kommen.

3.5 Anwendung und Evaluation

In wie weit die Anpassungen der Nutzung von Twitter als Vorhersageinstrument für politische Wahlen erfolgreich sind, um frühere Schwierigkeiten zu beheben, soll anhand einer Fallstudie untersucht werden. Dazu werden die Tweets in Deutschland innerhalb des Jahres 2013 bis zur Bundestagswahl genutzt. Daraufhin wird das Verfahren in Bezug auf die eigentlichen Ergebnisse der Bundestagswahl evaluiert. Auf der einen Seite werden dann die Tweets der letzten 4 Wochen vor der Wahl, abzüglich der letzten Woche (somit: 26. August 2013 - 15. September 2013) klassifiziert und die mit a korrigierte Stimmverteilung auf Twitter ermittelt. Die Vernachlässigung der Daten aus der letzten Woche vor der Wahl ergeben sich aus dem Anspruch, eine Woche vor einer Wahl eine Prognose für das Wahlergebnis abgeben zu können. Auf der anderen Seite wird bei den Wahlergebnissen der Anteil an den Zweitstimmen für die Parteien genutzt, wobei die Stimmen für die kleinen Parteien nicht berücksichtigt werden. Der sich ergebenden Fehler in Form des mean average error (MAE) wird daraufhin mit den Ergebnissen führender Forschungsgruppen für die

Vorhersage von Wahlergebnissen in Deutschland verglichen (Forsa, Allensbach, Amnid, Forschungsgruppe Wahlen, GMS, Infratest/dimap). Interessant ist auch die Frage, ob das knappe Scheitern der FDP und der AfD an der in Deutschland festgesetzten 5%-Hürde anhand der Prognose verhergesagt worden wäre. Die Auswahl der Parteien sieht dabei folgendermaßen aus: Union (CDU/CSU), SPD, Die Linke, Bündnis 90/Die Grünen, FDP, AfD und die Piratenpartei. Die Piratenpartei wird dabei explizit mit einbezogen, um die durch [JJS12] aufgeführte Problematik der jungen und internetaffinen Anhängerschaft dieser Partei einzubeziehen. Die AfD befindet sich in der Liste der berücksichtigten Parteien, da sie als neue Partei trotzdem als Kandidat für einen Eintritt in den Bundestag galt.

Zu klären gilt noch, welche Politiker der Parteien für den ersten Schritt des Verfahrens gewählt werden sollen. Für die ersten fünf genannten Parteien bietet es sich an, die Mitglieder des deutschen Bundestags der 17. Wahlperiode zu verwenden. Dies ist für die AfD und die Piratenpartei aber nicht anwendbar, da diese keine Mitglieder für den Bundestag stellen.

Zusätzlich zu der Evaluation des gesamten Verfahrens, bietet es sich an, den 1. Schritt (3.2) getrennt zu untersuchen. Die Genauigkeit des Classifiers lässt sich prüfen, indem die eigentlichen Trainingsdaten, bestehend aus den Tweets der Politiker, in eine neue Trainingsmenge und eine Validierungsmenge geteilt werden. Nachdem auf den Trainingsdaten ein neues Modell trainiert wird, werden die Tweets der Validierungsmenge klassifiziert. Da für diese Tweets die Parteizugehörigkeit bekannt ist, kann die Richtigkeit der Klassifikation ermittelt werden. Das beschriebene Verfahren heißt k -fache Kreuzvalidierung[Koh95]. Die Eingabedaten werden dabei in k möglichst gleichgroße Teilmengen unterteilt und in k Durchläufen wird jeweils eine Menge als Validierungsmenge gewählt und die restlichen Teilmengen bilden die Trainingsmenge. Die Güte des Klassifikators bildet sich dann aus dem Durchschnitt der Güte der Klassifikatoren in den k Durchläufen.

Literatur

- [BG13] BEWARDER, Manuel ; GRADL, Benedikt: Bundestag : Warum Twitter den Wahlsieg nicht vorhersagen kann. In: *DIE WELT* (2013), September. <http://www.welt.de/politik/deutschland/article119891348/Warum-Twitter-den-Wahlsieg-nicht-vorhersagen-kann.html>, Abruf: 2013-12-13
- [BMZ11] BOLLEN, Johan ; MAO, Huina ; ZENG, Xiaojun: Twitter mood predicts the stock market. In: *Journal of Computational Science* 2 (2011), Nr. 1, 1–8. <http://www.sciencedirect.com/science/article/pii/S187775031100007X>
- [dpa13] DPA: Studie: Viele Tweets bedeuten viele Wahlstimmen. In: *Die Zeit* (2013), August. <http://www.zeit.de/news/2013-08/13/internet-studie-viele-tweets-bedeuten-viele-wahlstimmen-13112255>, Abruf: 2013-12-18
- [GA12] GAYO-AVELLO, Daniel: I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper - A Balanced Survey on Election Prediction using Twitter Data. In: *arXiv preprint arXiv:1204.6441* (2012). <http://arxiv.org/abs/1204.6441>
- [JJS12] JUNGHERR, Andreas ; JÜRGENS, Pascal ; SCHOEN, Harald: Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment". In: *Social Science Computer Review* 30 (2012), Nr. 2, 229–234. <http://ssc.sagepub.com/content/30/2/229.short>
- [Jun13a] JUNGHERR, Andreas: Statistik-Zwischenruf: Twitter sagt nicht das Wahlergebnis vorher. In: *Spiegel Online* (2013), September. <http://www.spiegel.de/netzwelt/netzpolitik/statistik-zwischenruf-twitter-sagt-nicht-das-wahlergebnis-vorher-a-919915.html>, Abruf: 2013-12-13
- [Jun13b] JUNGHERR, Andreas: Tweets and Votes, a Special Relationship: The 2009 Federal Election in Germany. In: *Proceedings of the 2nd workshop on Politics, Elections and Data*, ACM, 2013, 5–14
- [Jup12] JUPP, Emily: Obama's victory tweet 'four more years' makes history. In: *The Independent* (2012), November. <http://www.independent.co.uk/news/world/americas/us-elections/obamas-victory-tweet-four-more-years-makes-history-8293642.html>, Abruf: 2014-01-06

- [Koh95] KOHAVI, Ron: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)* Bd. 14, 1995, 1137–1145
- [MLA⁺11] MISLOVE, Alan ; LEHMANN, Sune ; AHN, Yong-Yeol ; ONNELA, Jukka-Pekka ; ROSENQUIST, J. N.: Understanding the Demographics of Twitter Users. In: *Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011, 554–557
- [MMGA11] METAXAS, Panagiotis T. ; MUSTAFARAJ, Eni ; GAYO-AVELLO, Daniel: How (not) to predict elections. In: *Privacy, security, risk and trust (PASSAT), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (SocialCom)*, 2011, 165–171
- [NGTM13] NGUYEN, Dong ; GRAVEL, Rilana ; TRIESCHNIGG, Dolf ; MEDER, Theo: "How Old Do You Think I Am?": A Study of Language and Age in Twitter. In: *Seventh International AAAI Conference on Weblogs and Social Media*, 2013, 439–448
- [PD11] PAUL, Michael J. ; DREDZE, Mark: You Are What You Tweet: Analyzing Twitter for Public Health. In: *Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011, 265–272
- [SB12] SANG, Erik Tjong K. ; BOS, Johan: Predicting the 2011 dutch senate election results with twitter. In: *Proceedings of the Workshop on Semantic Analysis in Social Media*, 2012, 53–60
- [SOM10] SAKAKI, Takeshi ; OKAZAKI, Makoto ; MATSUO, Yutaka: Earthquake shakes Twitter users: real-time event detection by social sensors. In: *Proceedings of the 19th international conference on World wide web*, 2010, 851–860
- [TSSW10] TUMASJAN, Andranik ; SPRENGER, Timm O. ; SANDNER, Philipp G. ; WELPE, Isabell M.: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In: *Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010, 178–185