“It’s Modern Day Presidential! An Evaluation of the Effectiveness of Sentiment Analysis Tools on President Donald Trump’s Tweets”

Ann Masterton Perry[1][a], Terhi Nurmikko-Fuller[1][b] and Bernardo Pereira Nunes[2,3][c]

[1] *Australian National University, Canberra, Australia*

[2] *Pontifical Catholic University of Rio de Janeiro, Brazil*

[3] *Federal University of the State of Rio de Janeiro, Brazil*

Keywords: Sentiment Analysis, Social Media, Donald Trump, POTUS, Twitter, Sentiment Analysis Tools.

Abstract: This paper reports on an evaluation of five commonly used, lexicon-based sentiment analysis tools (MeaningCloud, ParallelDots, Repustate, RSentiment for R, SentiStrength), tested for accuracy against a collection of Trump’s tweets spanning from election day November 2016 to one year post inauguration (January 2018). Repustate was found to be the most accurate at 67.53%. Our preliminary analysis suggests that this percentage reflects Trump’s frequent inclusion of both positive and negative sentiments in a single tweet. Additionally to providing an evaluative comparison of sentiment analysis tools, a summary of shared features of a number of existing datasets containing Twitter content along with a comprehensive discussion is also provided.

1 INTRODUCTION

The President of the United States of America (POTUS), Donald Trump, is an active and unique social media user. As POTUS, his posts have real-world effects that go beyond those of other users, and his discourse is influencing politicians’ engagement, changing the acceptable language and expected behaviour of elected officials (Stolee and Canton, 2018).

But what is POTUS tweeting about? We hypothesise that sentiment analysis tools (developed for large datasets derived from a broad range of users) would become inaccurate at the granularity of a single contributor, where linguistic idiosyncrasies represent a greater percentage of the data. To verify this hypothesis, we evaluated the accuracy of five tools (MeaningCloud, ParallelDots, Repustate, RSentiment for R, SentiStrength) in classifying Trump’s tweets.

In this paper, we summarise on existing research in the sentiment analysis of politicians’ use of Twitter (Section 2); outline the ethical considerations of including deleted tweets (Section 3); describe our methodology (Section 4) and the dataset (Section 5); and provide a comparative evaluation of five sentiment analysis tools (Section 6). Finally, we conclude the paper with a discussion (Section 7).

[a] https://orcid.org/0000-0001-8283-2246
[b] https://orcid.org/0000-0002-0688-3006
[c] https://orcid.org/0000-0001-9764-9401

2 BACKGROUND

(Stieglitz and Dang-Xuan, 2013) (Larsson and Kalsnes, 2014) (Gainous and Wagner, 2014) (Hoffman et al., 2016) have examined elected officials’ daily engagement with the public and changes to it within an election context (Strandberg, 2012) (Graham et al., 2014). Trump’s use of social media has been examined previously (Francia, 2017) (Gross and Johnson, 2016) (Enli, 2017) (Oh and Kumar, 2017) (Ahmadian et al., 2016) (Stolee and Canton, 2018) and (Karpf, 2017) found that his use of Twitter has changed the way political campaigns are conducted. This includes altering previously accepted models of political discourse in the United States (Ott, 2016) (Auxier and Golbeck, 2017) and legitimizing the behaviour of other world leaders (McNair, 2018).

The first instance of real-time sentiment analysis for political events was during the 2012 Obama-Romney presidential cycle, when (Wang et al., 2012) analysed over 36 million tweets posted during the campaign and election. Other studies into politicians’ Twitter use have been undertaken by (Park et al., 2015) (Ahmed et al., 2016) and (Wang et al., 2016).

Tweets contain hashtags (e.g #TrumpsWallSongs), acronyms (“LOL”), emoticons (:)) and

Perry, A., Nurmikko-Fuller, T. and Nunes, B.
“It’s Modern Day Presidential! An Evaluation of the Effectiveness of Sentiment Analysis Tools on President Donald Trump’s Tweets”.
DOI: 10.5220/0007759306440651
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 644-651
ISBN: 978-989-758-372-8

emojis (😔), which are problematic for lexicon-based analysis tools (Davidov et al., 2010) (Kouloumpis et al., 2011), but (Pak and Paroubek, 2010) developed accurate models using a naïve Bayes classifier.

Sarcasm and irony also presents a challenge (Reyes et al., 2012). Whilst models that consider *word context* have been successful (achieving 65% accuracy (Mukherjee and Bala, 2017)), (Wang et al., 2012) argue there is much work to be done in developing models for their identification and classification.

Trump's tweets do not typically follow English syntax, containing incomplete conditional clauses, erratic punctuation, and the use of capitals for emphasis, demonstrating his difference from the Washington political elite. His willingness to immediately tweet his thoughts on breaking news appeals to his supporters, who are disillusioned with the current political landscape and, even if his statements later turn out to be incorrect, this spontaneity is seen as a measure of honesty and forthrightness (Stolee and Canton, 2018).

Lakoff[1] argues that a key to Trump's success is that he appears to be one step ahead of other politicians and the media in communicating his thoughts on current events, framing events in his own context and perspective. He suggests that Trump's tweets are so-called "trial balloons", deliberately designed to gauge public response with no intention of commitment to an underlying policy. Ongoing coverage by mainstream media then legitimizes them, manufacturing consensus online (Woolley and Guilbeault, 2017).

3 PRIVACY IN THE PUBLIC EYE

The debate over whether or not publicly available social media data[2,3] should be accessible for academic research without requiring informed consent from each individual user is on-going (Bonnilla and Rosa, 2015) (Webb et al., 2017) (boyd and Crawford, 2012) (Nunan and Yenicioglu, 2013) (Fiesler and Proferes, 2018). In this paper, we argue that the content posted from the @realDonaldTrump account is *intended for public consumption*, giving three reasons: i) whether as POTUS, a high profile businessman, or a reality television star (all highly visible public positions which he actively sought), Trump can have no expectation of obscurity, nor likely a desire for it (Stolee and Canton, 2018); ii) it is possible both Trump and members of his administration are responsible for tweet content (Auxier and Golbeck, 2017), thus, they do not represent the output of an *individual*, as they do an *institution*, and crucially; iii) current interpretation of constitutional law in the United States holds that post-election tweets from @realDonaldTrump constitute Presidential records, and require preservation under the Presidential Records Act of 1978, and part of the public record.

This distinction between public and private is important in the context of POTUS' deleted tweets. (Maddock et al., 2015) identify that not only do legal obligations require researchers to remove deleted content, but so too do ethical obligations, as the act of deleting a tweet indicates withdrawal of consent for its use in research. Twitter's Developer Agreement requires that "all reasonable efforts to delete or modify" deleted content as soon as possible, or within 24 hours after being asked to do so by Twitter or the user[4], but (Meeks, 2018) argues that researchers may be able to use these tweets if they have been sourced from a third party, such as Politiwoops[5], who have an agreement in place with Twitter to archive and publish deleted content. Furthermore, since content from the @realDonaldTrump is subject to preservation orders, all tweets, including any deleted ones, require preservation under the Presidential Records Act of 1978, and form part of the public record. We follow (Meeks, 2018), (Abramson, 2017), and (Dawsey and Bender, 2017), and have included deleted tweets – sourced by a third party (namely Politiwoops) – into our dataset.

4 METHODOLOGY

An initial survey of 64 existing tools led to the identification of five (MeaningCloud[6], ParallelDots[7], Repustate[8], RSentiment for R[9], and SentiStrength[10]) that met six selection criteria: i) ability to perform English language sentiment analysis on tweets, ii) are free to use, iii) do not require an existing application, iv) use a pre-built lexicon, v) are able to ingest the data collected for analysis, and vi) have no special computational infrastructure requirements (Table 1).

[1]https://georgelakoff.com/2017/03/07/trumps-twitter-distraction/

[2]https://twitter.com/en/privacy

[3]https://developer.twitter.com/en/developer-terms/agreement

[4]https://developer.twitter.com/en/developer-terms/policy

[5]https://projects.propublica.org/politwoops/

[6]http://www.meaningcloud.com

[7]http://www.paralleldots.com

[8]http://www.repustate.com

[9]https://cran.r-project.org/web/packages/RSentiment/index.html

[10]http://sentistrength.wlv.ac.uk/

Table 1: Tool selection criteria and required response.

Selection Criteria	Required
Suitable for English?	Yes
Suitable for ngram analysis?	Yes
Free to use?	Yes
API dependent?	No
Pre-built lexicon?	Yes
Able to ingest the dataset?	Yes
Special computational requirements?	No

Data was sourced from Politiwoops[11] and Factba.se[12]. They have a different collection rate, thus minimising the risk of missing or deleted tweets. Following data matching and a cross check against archives of deleted tweets from @realDonaldTrump, a consolidated dataset of 2,880 tweets was produced.

Manual sentiment annotation was completed by three coders on the crowdsourcing platform Figure Eight[13]. For quality assurance, we also manually annotated the tweets. Mismatches between the sentiment determined by the Figure Eight coders and us were referred to a third-party volunteer arbiter.

5 DATASET

5.1 Existing Twitter Datasets

A review of publicly available Twitter datasets (based on the work of (Saif et al., 2013) (Saif et al., 2012) (Saif et al., 2016), and confirmed and extended by (Symeonidis et al., 2018)) informed the development of the custom dataset used in the evaluation of the tools. The focus was on attributes such as size, creation method, and classifier categories. A total of 10 datasets (six of which were publicly available at the time of writing) were identified. They were reviewed to assess size, classifiers used (e.g. positive, negative or neutral) and the creation workflow (see Table 2).

The Debate08 or Obama-McCain Debate (OMD) dataset holds 3,238 tweets (Shamma et al., 2009)[15] that were manually annotated by three coders using classifiers for *positive*, *negative*, *mixed* or *other*. It has been utilized for testing supervised learning methods by (Saif et al., 2012) and (Hu et al., 2013).

The Stanford Semantic Twitter Sentiment (STS) Datasets were among the first of their kind. Initially built by (Go et al., 2009) to support a project report, the dataset contains training and testing corpora, both using three classifications (*positive*, *negative* and *neutral*). The 1.6 million tweets of the STS-Training dataset were automatically annotated, whilst the smaller STS-Test dataset was hand-coded.

The Health Care Reform (HCR) dataset[16] comprises of 2,516 tweets that contain the #hcr hashtag, which relates to the efforts of the Obama administration to introduce the Affordable Care Act. A subset of this corpus was manually annotated by the authors with five labels (*positive*, *negative*, *neutral*, *irrelevant*, and *other*). It is comprised of three sub-corpora for use in model development, evaluation, and training.

The STS-Gold Dataset (constructed by (Saif et al., 2013) from the Stanford datasets) comprises of 2,142 tweets, classified as *positive*, *negative*, *neutral*, *mixed* and *other*. It was hand-coded by three annotators.

The Sentiment Strength[17] (SS) Twitter Dataset (SS-Twitter) was developed by (Thelwall et al., 2010) to evaluate the effectiveness of their lexicon-based tool, SentiStrength. It was hand-annotated by three annotators to assign a Likert-type numerical value (both a positive (1 [not positive] to 5 [extremely positive]) and a negative (-1 [not negative] to -5 [extremely negative]) one) to each of the 4,242 tweets.

5.2 Dataset Used for Tool Evaluation

Our dataset was created by combining tweets from the Trump Twitter Archive[18], and those directly extracted from Twitter using a third party tool, FireAnt[19]. There were two challenges to the process of data selection and acquisition: i) access via the public API is constrained to the last seven days, and ii) deleted tweets.

Our dataset comprises of 2,981 tweets from election day 2016 (8 November) to one year post the Presidential inauguration (20 January, 2017). This period was chosen as it is slightly longer than a full calendar year of events in a Presidential diary. This provided the opportunity to include regular events including international assemblies such as the Group of Seven (G7) summit, holidays and recurrent natural cycles such as the US hurricane season. The period also covers a broad range of unanticipated national and international events, such as domestic protests

[11] https://projects.propublica.org/politwoops/user/realDonaldTrump

[12] https://factba.se/topic/deleted-tweets

[13] https://www.figure-eight.com, formerly known as Crowdflower

[15] Posted during a 2008 presidential debate between the President Obama and challenger Senator McCain

[16] https://bitbucket.org/speriosu/updown/src/1deb8fe45f603a61d723cc9b987ae4f36cbe6b16/data/hcr/?at=default

[17] http://sentistrength.wlv.ac.uk/#About

[18] http://www.trumptwitterarchive.com/

[19] http://www.laurenceanthony.net/software/fireant/

Table 2: Twitter dataset comparison.

Dataset	Total Tweets	Coding	Coders	Use	Classifiers used					
					positive	neutral	negative	mixed	other	relevant
STS-Test	489	manual	3+	test	yes	yes	yes	no	no	yes
STS-Gold	2,124	manual	3	test	yes	yes	yes	yes	yes	yes
STS-Training	1.6M	auto	n/a	train	yes	yes	yes	no	no	yes
HCR	2,516	manual	?	all[14]	yes	yes	yes	no	yes	no
Debate08 (OMD)	3,238	manual	3+	test	yes	no	yes	yes	yes	yes
SS-Twitter	4,242	manual	3+	test	yes	yes	yes	no	no	yes

(Chenoweth and Pressman, 2017) and the testing of nuclear weapons by North Korea (Fifield, 2017).

6 TOOL EVALUATION

6.1 T1: MeaningCloud

MeaningCloud[20] provides topic extraction, text classification, sentiment analysis and summarization through Excel add-ons, plug-ins, and cloud-based APIs. It uses a pre-built dictionary to determine sentiment based on (uniquely amongst the evaluated tools) four possibilities: *positive*, *negative*, *neutral*, or *none*, but was unable to classify 13.50% of the data.

MeaningCloud was the second most accurate tool in correctly identifying *positive* tweets, with a success of 72.15% (see T1 in Table 3). It accurately determined *negative* tweets 50.54% of the time. Many of those mis-identified as *positive* relate to attacks on main stream media ("Drain the Swamp should be changed to Drain the Sewer - it's actually much worse than anyone ever thought, and it begins with the Fake News!"[21]) or the investigation into Russian influence on the election ("This is the single greatest witch hunt of a politician in American history!"[22]). *Positive* tweets mis-classified as *negative* have mixed language, mention the economy or the inauguration.

Most of the *neutral* tweets that were mis-identified as *negative* express condolences or remembrance ("National Pearl Harbor Remembrance Day - "A day that will live in infamy!" December 7, 1941"[23]), while those *neutral* tweets that were misidentified as *positive* were concerned with natural disasters.

[20] http://www.meaningcloud.com

[21] https://twitter.com/realdonaldtrump/status/889435104841523201

[22] https://twitter.com/realdonaldtrump/status/865173176854204416

[23] https://twitter.com/realdonaldtrump/status/938786402992578560

Table 3: Accuracy of all five sentiment analysis tools.

	Negative	**Neutral**	**Positive**	**Total**
T1	50.54%	5.00%	72.15%	62.95%
T2	64.70%	52.50%	64.50%	64.41%
T3	64.07%	35.00%	70.48%	67.53%
T4	37.66%	45.00%	73.88%	59.62%
T5	60.07%	47.50%	58.52%	58.96%

6.2 T2: ParallelDots

ParallelDots[24] provides machine learning services for text analytics. They offer Microsoft Excel, Google Sheets add-ins and cloud-based APIs. These enable use of functionality including keyword extractions, named entity recognition and sentiment and emotion analysis, using a pre-built lexicon to classify each tweet as either *positive*, *negative* or *neutral*.

Overall accuracy was 64.41%, correctly identifying 64.50% of *positive*, 64.70% of *negative* and 52.50% of *neutral* tweets. Self-promotive tweets proved an issue, 41.65% of which were misidentified.

There was some consistency among the categories it misidentified: tweets regarding players protesting by kneeling for the national anthem as *positive* even though they are negative in sentiment ("The NFL is now thinking about a new idea - keeping teams in the Locker Room during the National Anthem next season. That's almost as bad as kneeling! When will the highly paid Commissioner finally get tough and smart? This issue is killing your league!....."[25]), as were tweets regarding immigration and the border wall ("The judge opens up our country to potential terrorists and others that do not have our best interests at heart. Bad people are very happy!"[26]).

[24] http://www.paralleldots.com

[25] https://twitter.com/realdonaldtrump/status/933285973277868032

[26] https://twitter.com/realdonaldtrump/status/828042506851934209

6.3 T3: Repustate

Repustate[27] provides server-based software, APIs and added functionality to Microsoft Excel. Results are provided numerically from 1 to -1: the greater the number, the more *positive* the tweet, the lower a negative number, the more *negative*. A score of zero indicates a *neutral* tweet.

This tool had the highest overall accuracy of 67.53%. It ranked third in identifying *positive* tweets (70.10%). Self-promotional tweets were the most challenging to identify, mis-classified in 32.65% of cases. It was the second most accurate in identifying *negative* tweets (64.07%). Close reading did not identify the use of positive language, nor tweets that are purely informational in those categorized *neutral*. Entries such as “Obama Administration official said they “choked” when it came to acting on Russian meddling of election. They didn’t want to hurt Hillary?”[28] ?” are classified as *positive*, whilst “The judge opens up our country to potential terrorists and others that do not have our best interests at heart. Bad people are very happy!” [29] to be identified as *neutral*, possibly since they combine both negative and positive sentiments that effectively cancel each other out.

Only 35.00% of *neutral* tweets were classified correctly. These tweets included words such as “harm”, “destructive” and “terrorism” (“Today, I signed the Global War on Terrorism War Memorial Act (#HR873.) The bill authorizes....cont https://t.co/c3zIkdtowc https://t.co/re6n0MS0cj”[30]), illustrating that using only words tagged with predetermined sentiment can be problematic.

6.4 T4: RSentiment Package for R

RSentiment[31] is an Open Source package for use with R. It provides a range of queries for sentiment determination through a pre-built lexicon and can also ingest a custom list of single words. It uses Part-of-Speech (POS) tagging to identify nouns, verbs, adjective, adverbs, etc., inferring context from the order of these words and incorporating this in the sentiment score. Additional calculations are performed to identify negation and sarcasm within the sentence.

For this evaluation the calculate sentiment function was used. This provides results across five categories: *very positive*, *positive*, *neutral*, *negative* and *very negative* (Bose, 2018): these results were converted to three categories (*positive*, *negative* and *neutral*) for consistency with the other evaluated tools.

The heavy draw on system resources complicated the use of this tool, with memory errors consistently encountered when processing more than 200 tweets at a time. The dataset had to be split into 15 separate files with a maximum of 200 entries. Each file was imported, analysed, and exported separately, and results consolidated on completion.

This tool had the second most *inaccurate* result with a total accuracy of 59.62%. It was most inaccurate in identifying *negative* tweets (only 37.66% were correctly classified). Analysis of the words used in these tweets does not provide a cause for this, given the tool identified tweets such as “The @nytimes sent a letter to their subscribers apologizing for their BAD coverage of me. I wonder if it will change - doubt it?”[32] as *positive*, and “Just tried watching Saturday Night Live - unwatchable! Totally biased, not funny and the Baldwin impersonation just can’t get any worse. Sad”[33] as *neutral*.

RSentiment was unable to correctly identify a majority of neutral tweets scoring 45.00%. As with previous tools, it is easy to identify words such as “holocaust” and “destructive” that cause negative classifications, however, it is unclear which words in tweets such as “RT @NWSHouston: Historic flooding is still ongoing across the area. If evacuated, please DO NOT return home until authorities indicate it i...”[34] would lead to a *positive* classification.

RSentiment had an accuracy of 73.88%. Self-promotive tweets were the most difficult to classify, accounting for more than half of the misidentified tweets. It is clear how particular words have driven a *negative* classification in some instances (“Despite the phony Witch Hunt going on in America, the economic & jobs numbers are great. Regulations way down, jobs and enthusiasm way up!”[35]),but not in others, e.g.“A great great honor to welcome & recognize the National Teacher of the Year, as well as the Teacher of the Year fro… https://t.co/pUGl7RDoVX”[36]. as they contain several terms generally thought of as positive.

[27] http://www.repustate.com
[28] https://twitter.com/realdonaldtrump/status/878715504063643648
[29] https://twitter.com/realdonaldtrump/status/828042506851934209
[30] https://twitter.com/realdonaldtrump/status/898718902200418306
[31] https://cran.r-project.org/web/packages/RSentiment/index.html
[32] https://twitter.com/realdonaldtrump/status/797812048805695488
[33] https://twitter.com/realdonaldtrump/status/805278955150471168
[34] https://twitter.com/realdonaldtrump/status/902491685720076288
[35] https://twitter.com/realdonaldtrump/status/875698062030778368
[36] https://twitter.com/realdonaldtrump/status/

6.5 T5: SentiStrength

SentiStrength[37] is an opinion-mining software (SentiStrength, 2018). It uses a pre-built dictionary of words tagged *positive*, *negative* and *neutral* to determine sentiment. There are four options for output:

- Dual: where both a positive (from 1 - 5) and negative (from -1 to -5) score is given for each tweet
- Binary: positive and negative
- Trinary: positive, negative and neutral
- Scale: sentiment is rated -4 to +4, representing negative to positive results respectively.

The trinary (positive, negative and neutral) setting was initially chosen for analysis to maintain consistency. This setting did not work as expected with results being presented as dual. Testing of the Binary and Scale outputs revealed the same issue. To convert the dual score to comparable sentiment categories, the positive and negative scores were added together, and a classification generated. For example, if a tweet scored +3 for positive sentiment, and -1 for negative sentiment, it's overall score was calculated as +2, giving it an overall positive classification.

The relative ease-of-use and fast processing time (<5 seconds) of this tool are negated by it being the least accurate for overall tweet classification, at 58.96%. It was the least accurate of all tools in identifying *positive* tweets (58.52%), mis-identifying self-promotional tweets as *positive*. Topics such as foreign affairs and the economy were frequently misidentified, but in no observable pattern.

Tweets such as "China has been taking out massive amounts of money & wealth from the U.S. in totally one-sided trade, but won't help with North Korea. Nice!"[38] using mixed language and terms were mis-classified as positive. It is not clear why some very negative records were classified as neutral ("The Fake News media is officially out of control. They will do or say anything in order to get attention - never been a time like this!"[39]).

Sentistrength ranked second for identifying neutral tweets at 47.05% accuracy. Similar to other tools, it could not identify neutral tweets related to remembrance or disasters where typically negative words such as battlefield or storm were used, and misclassified some records as positive when terms including "strengthens", "bless" or "pioneer" were present.

857360510534209536

[37] http://sentistrength.wlv.ac.uk/

[38] https://twitter.com/realdonaldtrump/status/816068355555815424

[39] https://twitter.com/realdonaldtrump/status/86008733451941478

7 CONCLUSION

The tweets of POTUS Donald Trump have provided his social media audience with rich and varied materials. The first stage to understanding the true nature of this content has been the evaluation of five existing sentiment analysis tools, selected from an original list of over 60 possibilities. The results from our evaluative testing show that MeaningCloud accurately classified 62.95% of the tweets in our purpose-built dataset; ParallelDots was 64.41% accurate; Repustate was 67.53%; RSentiment was 59.62% accurate; and SentiStrength was 58.96% accurate (Table 3).

There is consistency in the results: MeaningCloud, Repustate and RSentiment were all more accurate at determining positive sentiment, and all five tools were least able to identify neutrality. Tweets of self-promotion were the most problematic for all tools, possibly since these typically contain a high level of mixed language, often attacking an opponent then promoting Trump or his administration's achievements. Neutral tweets were often misclassified due to their inclusion of language related to death or natural disasters. Determining sentiment using a pre-tagged dictionary is problematic, as the tools fail to recognise the context in which these terms are used, and cannot be guaranteed to distinguish between a statement of fact and a negative expression.

In addition to providing an evaluative comparison of the five tools, the paper provides a summary of shared features of a number of existing datasets containing Twitter content, which are freely accessible. The project as it currently stands has relied heavily on such a third-party archive of tweets as well as a third-party tool for gathering tweets, but there has been little scope for critically evaluating either the third-party dataset or the tweet-gathering tool, beyond expressing a concern regarding missing or deleted tweets, and noting an issue with timezones: to simplify the task, we made a baseline assumption that all of Trump's tweets were posted in Eastern standard time, although the documentation for the third-party archive does include the caveat that they cannot be certain when he was tweeting from other parts of the country or world.

Another assumption made in the process of this analysis that the @realDonaldTrump Twitter account captures the idiosyncratic voice of a single user. It is however possible, that both Trump and his political team use the account, with some evidence that, for example, prior to the election, there was a difference in the voice dependent on the type of equipment used (iPhone or Android); after the election, we could assume that tweets posted between the hours of 10pm - 9am are likely Trump personally, but cannot assert the same with equal confidence during office hours, for example. Future work will concentrate on analyses, which may be in a position to use natural language processing and sentiment analysis to investigate whether one or more author-voices can be detected in the data.

ACKNOWLEDGEMENTS

The work carried out for this paper formed part of an unpublished Master's thesis at the Australian National University, which, in turn, builds on an unpublished preliminary investigation submitted for the *SOCR8006 Online Research Methods* course taught by Associate Professor Robert Ackland at the Australian National University. The authors would like to acknowledge and thank all their colleagues who have contributed to any and all of these pieces, including Professor Les Carr, University of Southampton, and Dr Jenny Davis, Australian National University, who acted as examiners for the thesis.

REFERENCES

Abramson, A. (2017). *President Trump Just Deleted Tweets About Luther Strange. Is He Allowed to Do That?* http://time.com/4958877/donald-trump-deletes-tweets-legal-implications/.

Ahmadian, S., Azashahi, S., and Palhus, D. L. (2016). Explaining donald trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences*, 107:49–53.

Ahmed, S., Jaidka, K., and Cho, J. (2016). The 2014 indian elections on twitter: A comparison of campaign strategies of political parties. *Telematics and Informatics*, 33:1071–1087.

Auxier, B. and Golbeck, J. (2017). The president on twitter: A characterization study of @realdonaldtrump. In Ciampaglia, G., Mashhadi, A., and Yasseri, T., editors, *Social Informatics. SocInfo 2017. Lecture Notes in Computer Science*, volume 10539, pages 377–390.

Bonnilla, Y. and Rosa, J. (2015). #fergusonl digital protest, hashtag ethnography, and the racial politics of social media in the united states. *American Ethnologist*, 42(1):4–17.

Bose, S. (2018). Rsentiment: Analyse sentiment of enlgish sentences. r package version 2.2.2.

boyd, d. and Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5):662–679.

Chenoweth, E. and Pressman, J. (2017). *Last month, 83% of U.S. protests were against Trump.* https://www.washingtonpost.com/news/monkey-cage/wp/2017/09/25/charlottesville-and-its-aftermath-brought-out-many-protesters-in-august-but-still-more-were-against-trump-and-his-policies/?noredirect=on&utm_term=.5ea64cd2742f.

Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters*, pages 241–249. Association for Computational Linguistics.

Dawsey, J. and Bender, B. (2017). National archives warned trump white house to preserve documents.

Enli, G. (2017). Twitter as arena for the authentic outsider: Exploring the social media campaigns of trump and clinton in the 2016 us presidential election. *European Journal of Communication*, 32:50–61.

Fiesler, C. and Proferes, N. (2018). "participant" perceptions of twitter research ethics. *Social Media+ Society*, 4(1):2056305118763366.

Fifield, A. (2017). *In latest test, North Korea detonates its most powerful nuclear device yet.* https://www.washingtonpost.com/world/north-korea-apparently-conducts-another-nuclear-test-south-korea-says/2017/09/03/7bce3ff6-905b-11e7-8df5-c2e5cf46c1e2_story.html?utm_term=.462a0a2b8411.

Francia, P. L. (2017). Free media and twitter in the 2016 presidential election: The unconvensional campaign of donald trump. *Social Science Computer Review*, page 0894439317730302.

Gainous, J. and Wagner, K. M. (2014). *Politics, Tweeting to Power: The Social Media Revoluion in American Politics*. Oxford University Press, New York.

Go, a., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).

Graham, T., Jackson, D., and Broersma, M. (2014). New platform, old habits? candidates' use of twitter during the 2010 british and dutch general election campaigns. *New Media and Society*, 18(5):765–783.

Gross, J. H. and Johnson, K. T. (2016). Twitter taunts and tirades: Negative campaigning in the age of trump. *PS: Political Science & Politics*, 49(4 (Elections in Focus)):748–754.

Hoffman, C. P., Suphan, A., and Meckel, M. (2016). The impact of use motives on politicians' social media adoption. *Journal of Information Technoloyg and Politics*, 13(3).

Hu, X., Tang, L., Tang, J., and Liu, H. (2013). Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 537–546. ACM.

Karpf, D. (2017). Digital politics after trump. *Annals of the International Communication Association*, 41(2):198–207.

Kouloumpis, E., Wilson, T., and Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Icswm*, volume 11, page 164.

Larsson, A. O. and Kalsnes, B. (2014). 'of course we are on facebook': Use and non-use of social media among swedish and norwegian politicians. *European Journal of Communication*, 29(6):653–667.

Maddock, J., Starbird, K., and Mason, R. M. (2015). Using historical twitter data for research: Ethical challenges of tweet deletions. In *CSCW 2015 Workshop on Ethics for Studying Sociotechnical Systems in a Big Data World. ACM*.

McNair, B. (2018). From control to chaos, and back again. *Journalism Studies*, 19(4):499–511.

Meeks, L. (2018). Tweeted, deleted: theoretical, methodological, and ethical considerations for examining politicians' deleted tweets. *Information, Communication & Society*, 21(1):1–13.

Mukherjee, S. and Bala, P. K. (2017). Sarcasm detection in microblogs using naïve bayes and fuzzy clustering. *Technology in Society*, 48:19–27.

Nunan, D. and Yenicioglu, B. (2013). Informed, uninformed and participative consent in social media research. *International Journal of Market Research*, 55(6):791–808.

Oh, C. and Kumar, S. (2017). How trump won: The role of social media sentiment in political elections. In *Pacific Asia Conference on Information Systems (PACIS 2017)*, volume 17, page 48.

Ott, B. L. (2016). The age of twitter: Donald j. trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1):59–68.

Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10 No 2010, pages 1320–1326.

Park, S., Park, J. Y., Lim, Y. S., and Park, H. W. (2015). Expanding the presidential debate by tweeting: The 2012 presidential debate in south korea. *Telematics and Informatics*.

Reyes, A., Rosso, P., and Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data and Knowledge Engineering*, 74:1–12.

Saif, H., Fernandez, M., He, Y., and Alani, H. (2013). Evaluation datasets for twitter sentiment analysis: A survey and a new dataset, the sts-gold. In *1st Intl Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013)*, pages 9–21.

Saif, H., He, Y., and Alani, H. (2012). Semantic sentiment analysis of twitter. In *Proceedings of the 11th Intl conference on The Semantic Web*, pages 508–524.

Saif, H., He, Y., Fernandez, M., and Harith, A. (2016). Contextual semantics for sentiment analysis of twitter. *Information Processing and Management*, 52(1):5–19.

Shamma, D. A., Kennedy, L., and Churchill, E. F. (2009). *Obama McCain Debate Dataset*. https://bitbucket.org/speriosu/updown/wiki/Getting_Started.

Stieglitz, S. and Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4):1277–1291.

Stolee, G. and Canton, S. (2018). Twitter, trump, and the base: A shift to a new form of presidential talk? *Signs and Society*, 6(1):147–165.

Strandberg, K. (2012). A social media revolution or just a case of history repeating itself? the use of social media in the 2011 finnish parliamentary elections. *New Media and Society*, 15(8):1329–1347.

Symeonidis, S., Effrosynidis, D., and Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Systems With Applications*, 110:298–310.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.

Wang, H., Can, Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *ACL 2012 System Demonstrations*, pages 115–120.

Wang, Y., Li, Y., and Luo, J. (2016). Deciphering the 2016 us presidential campaign in the twitter sphere: A comparison of the trumpists and clintonists. In *ICWSM*, pages 723–726.

Webb, H., Jirotka, M., Stahl, B. C., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O., and Burnap, P. (2017). The ethical challenges of publishing twitter data for research dissemination. In *Proc of the ACM on Web Science Conference*, pages 339–348. ACM.

Woolley, S. C. and Guilbeault, D. R. (2017). Computational propaganda in the united states of america: Manufacturing consensus online. *Computational Propaganda Research Project*, page 22.