Erratum: Employing sentiment analysis for gauging perceptions of minorities in multicultural societies: An analysis of Twitter feeds on the Afrikaner community of Orania in South Africa

CrossMark
click for updates

Authors:
Eduan Kotzé[1]
Burgert Senekal[2]

Affiliations:
[1]Department of Computer Science and Informatics, University of the Free State, South Africa

[2]Unit for Language Facilitation and Empowerment, University of the Free State, South Africa

Corresponding author:
Eduan Kotzé,
kotzeje@ufs.ac.za

In the version of this article initially published, the text entries in Eqn 1 and Eqn 4 were mistakenly misspelled which presented it in an illegible format. The actual values for Eqn 1 and Eqn 4 are updated and presented here:

$$Precision\ (p) = \frac{TP}{TP + FP} \qquad \text{[Eqn 1]}$$

$$F - measure = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \qquad \text{[Eqn 4]}$$

These corrections do not alter the study's findings of significance or overall interpretation of the study results. The errors have been corrected in the PDF version of the article. The publisher apologises for any inconvenience caused.

AOSIS

Employing sentiment analysis for gauging perceptions of minorities in multicultural societies: An analysis of Twitter feeds on the Afrikaner community of Orania in South Africa

CrossMark
click for updates

Authors:
Eduan Kotzé[1] ⓘ
Burgert Senekal[2] ⓘ

Affiliations:
[1]Department of Computer Science and Informatics, University of the Free State, South Africa

[2]Unit for Language Facilitation and Empowerment, University of the Free State, South Africa

Corresponding author:
Eduan Kotzé,
kotzeje@ufs.ac.za

Read online:
Scan this QR code with your smart phone or mobile device to read online.

South Africa is well known as a country characterised by racial and ethnic divisions, particularly for the divisions and conflicts between the white population and black population. This study uses the Twitter platform to analyse the discourse around the controversial town of Orania, a minority Afrikaner community that aims to preserve their Afrikaner culture. In doing so, we make use of sentiment analysis, a subfield of natural language processing (NLP). We follow a lexicon-based approach using four different publicly available data sets to test how the discourse around this minority community can be analysed. We show, based on the discourse on Orania on Twitter, that (1) Orania is mostly depicted in a negative light, (2) Orania is mostly seen as a racist community, and (3) Orania is often mentioned in reference to other issues that affect Afrikaners directly, such as farm attacks, first language education and land expropriation without compensation. Our study also shows that using lexicons as a sentiment analysis technique was not sufficient in the automatic detection of abusive language, but rather the sentiment of the tweet. Suggestions are made for further research that focuses on the automatic detection of abusive language online.

Introduction

South Africa is a diverse country renowned for its racial and cultural tensions. In recent years, commentators such as Khoza (2017b), Roets (2017), Brink and Mulder (2017) and Steward (2016) have noted an increase in racial tensions as witnessed on social media platforms. Social media platforms are a group of internet-based applications that allow for the creation and exchange of data that are generated by people. Gundecha and Liu (2012) describe the different types of social media applications such as online social networking (Facebook, MySpace, LinkedIn), blogs (Engadget), microblogging (Twitter, Tumblr, Plurk), social news (Digg, Reddit), media sharing (YouTube, Flickr) and wikis (Wikipedia, Wikitravel, Wikihow). The most important social media platforms used to create and exchange user-generated content (UGC) are online social networking, microblogging and blogs (Stieglitz & Dang-Xuan 2013). One of the most prominent microblogging platforms is Twitter.com (http://www.twitter.com/), which has grown steadily in its adoption as the main microblogging platform in South Africa over the last few years. In a recent survey, World Wide Worx (2016) reports that Twitter is used by approximately 7.7 million people in South Africa, making it the third most used social media platform after YouTube (8.74 million) and Facebook (14 million). Some recent important debates under the hashtags #FeesMustFall and #StateCapture have also drawn much attention to how the general public can voice their opinions using social media (Findlay 2015), and one could add the conflict around the hashtag #WinnieMandela as a more recent example of how ethnic tensions play out on social media.

Sentiment analysis, a widely adopted big data analytics technique, is often used to mine customers' views and opinions from online social media platforms (Pang & Lee 2008). Sentiment analysis is a growing focus area of natural language processing (NLP) used to determine whether a text, or part of it, is subjective or not, and if subjective, whether it expresses a positive, negative or neutral view (Taboada 2016). Sentiment analysis of microblogging data, such as Twitter, has attracted much attention in recent years, both in industry and in academia. The reasons are mainly because of the rapid growth in Twitter's popularity as a platform for people to express their opinions and attitudes towards topics of interest. Not only is Twitter popular, the platform also contains an enormous volume of text as well as links to external media, including websites that are visible to subscribed users to that service (Pak & Paroubek 2010). This makes it an

invaluable research resource for sentiment analysis, a field associated with extracting sentiments (or opinions) from unstructured text (such as Twitter messages).

The purpose of this study is to investigate the discourse surrounding the community of Orania by using the sentiment analysis on tweets and whether this approach is effective at gauging the public perception about a minority community. The rest of this article is organised as follows. Firstly, we provide some background to the establishment of Orania. Thereafter, we describe business intelligence, sentiment analysis and related work. We then include an outline of the methodology and the data used for analysis, and highlight the results. After this section, we discuss the findings and whether it is applicable in gauging sentiment for a minority community.

Background to the establishment of Orania

Since united by the British Empire in 1910 when the Union of South Africa was established, South Africa has struggled with accommodating its various ethnic groups. Although best known for the conflict between black population and white population because of apartheid, these groups are themselves also diverse and have often been in conflict. One solution to the problem of handling such a diverse population, and in following the European example (see Muller 2008), is to divide the country so that each population can achieve self-rule while still maintaining economic and other ties. Muller (2008:27) argues that ethnic conflict was one of the primary causes of the two World Wars and that aligning national borders with populations was proposed by Winston Churchill, Franklin Roosevelt and Joseph Stalin as 'a prerequisite to a stable postwar order'. Muller (2008) also quotes Churchill from a speech to the British parliament in December 1944, where he referred to the forced resettlement of populations:

> Expulsion is the method which, so far as we have been able to see, will be the most satisfactory and lasting. There will be no mixture of populations to cause endless trouble. … A clean sweep will be made. (p. 27)

This was the original notion behind apartheid, which was formally introduced in 1948: Each population would gain independence and self-rule in their own geographic area. Prime Minister Verwoerd, for instance phrased this view clearly on 20 May 1959: 'Die een beginsel is die vrymaking van die Bantoe: die ander beginsel is die vrymaking van die blanke' [The one principle is the liberation of the Bantu: the other principle is the liberation of whites] (Pelzer 1966:275). In other words, no population group would rule over another; each would follow the European example where nations ruled themselves in their own geographic area. To do this, attempts were made to expand the existing homelands and grant them independence, with Transkei being the first to gain independence in 1976 and Bophuthatswana following in 1977. However, by this time it was already apparent that the homelands were not economically or politically viable, and

alternative forms of partition were debated by, for instance Tiryakian (1967), Sulzberger (1977), Von der Ropp (see Blenck & Von der Ropp 1977; Von der Ropp 1979; 1981; Von der Ropp & Blenck 1976), Lambsdorff (1986) and Pabst (1996) (see also Geldenhuys 1981:55). Von der Ropp, for instance proposed that South Africa would be divided into a white part and black part, which would allow each part access to mines, harbours and large metropolitan areas that would give people access to land and economic resources. Partition in the South African case did not and has not gained favour, neither nationally nor internationally, and in the run up to South Africa's first inclusive election in 1994, the leadership of most of these communities decided to opt for an inclusive 'Rainbow Nation' where all would integrate and work and live side by side.

One of the communities that challenged this integrative solution was the Afrikaner Nationalist community. Fearing that majority rule would eradicate minority rights, languages and communities (the Afrikaner currently comprises 59.1% of the white population, which in turn comprises around 8% of the total South African population), it was proposed that the Afrikaner establish a *volkstaat* (a country for the Afrikaner) where they would rule themselves (Hagen 2013; Pienaar 2007; Schönteich & Boshoff 2003). The idea of violent cessation was discussed in the early 1990s, but given the cost of a civil war, the leading organisation that proposed this solution, the Freedom Front, reached a settlement with the African National Congress (ANC) that led to the inclusion of Article 235 in the new South African constitution, which guarantees minorities' right to self-determination. No *volkstaat* was however established.

The Afrikaner Vryheidstigting (Afrikaner Freedom Foundation), founded on 21 March 1988 and currently known as the Orania Beweging (Orania Movement), however sought a practical rather than political solution. In 1991, they bought the abandoned town of Orania in the Northern Cape with the goal of establishing an Afrikaner community here. A few hundred Afrikaners moved here, and although initial progress was slow, in recent years this community has grown to around 1400 currently. Orania now has its own bank, the Orania Spaar-en Krediet Koöperatief (Orania Savings and Credit Cooperative); uses its own 'currency', the Ora (although tied to the South African rand); has a fast-growing economy; and exports products globally (see De Beer 2006; Hagen 2013; Kotze 2003; Labuschagne 2008; Pienaar 2007; Steyn 2005).

Part of the reason for Orania's recent growth has been the Afrikaner's increasing sense of marginalisation, as studied, for instance, by Hermann (2006). Affirmative action, the loss of first language education, negligible political power, crime, farm attacks and threats of violence by black political leaders have all contributed to many Afrikaners questioning the viability of the Rainbow Nation. For instance, the leader of the third-largest political party, the Economic Freedom Fighters (EFF), Julius Malema, has repeatedly called for the dispossession of 'whites'' property and urged his followers to kill 'Boers' (Afrikaners). In 2017, Brink and Mulder (2017)

compiled a report that shows numerous government officials calling for violence against Afrikaners, while commentators such as Khoza (2017b) and Steward (2016) also note the increasing amount of hate speech directed towards whites on social media. At the time of writing, the South African government is in the process of evaluating how to expropriate land without compensation, and the discourse around the subject is often phrased in ethnic terms: black people are the 'original' owners and white people 'stole' the land and should therefore be dispossessed (see, e.g. Eloff 2017; Osborne 2018). Importantly, a report by the South African Institute on Race Relations, based on a nationwide study, notes that:

> … 61% of black respondents now agree that South Africa is a country for blacks rather than whites, while only 38% disagree. This suggests that ANC and EFF rhetoric castigating whites and demanding a major shift in the ownership and management of the economy may be having significant impact on black opinion (Jeffery 2018).

From its inception, Orania has been the target of fierce criticism. Orania is often depicted in the media as a racist town, a leftover of Apartheid populated by white people who refuse to abandon their prejudices (see, e.g. Khan 2014; McNally 2010; Ngugi 2017). Focusing on Afrikaner culture, Orania does not exclude anyone based on race, but in reality, the Afrikaner – as a descendant of European settlers since 1652 – is a white community and no black people have settled in Orania. This has led to Orania becoming a synonym of racism.

Literature review
Business intelligence and sentiment analysis

Business intelligence and analytics (BI & A) is becoming increasingly important in analysing UGC, which includes sentiments, images and videos using big data analytics (Chen, Chiang & Storey 2012). Effective BI & A can be used to improve a firm or an organisation's decision-making capabilities. Other uses include improving operations, reducing marketing costs or simply obtaining a better understanding of customer preferences and opinions (Wixom & Watson 2010).

One such an example is Twitter Sentiment Analysis (TSA), where text mining techniques are used to mine messages posted on Twitter. Twitter, a microblogging platform, allows users to share short messages, links to other websites, images or videos. The message is written by one person and read by a number of individuals, called followers. Most messages also contain hashtags, which in turn are used to indicate the relevance of a tweet to a certain topic. These hashtags are created using the # character, followed by the name of topic (#topic). Twitter Sentiment Analysis tends to focus on the sentiment identification or sentiment classification of individual Twitter messages, called *tweets*. Generally, two main approaches are followed for tweet-level sentiment detection: machine learning and lexicon-based.

The machine learning approach uses a sentiment classifier to determine the polarity of new texts (document, sentence or phrase). This process is referred to as supervised learning and requires training data to teach the sentiment classifier characteristics which distinguish a negative sentiment from a positive one (Pang, Lee & Vaithyanathan 2002). The training data are usually labelled according to the tweet's polarity (positive, negative and neutral) and can be inferred using hashtags and emoticons (Go, Bhayani & Huang 2009), or by means of consensus using results from Twitter sentiment websites (Barbosa & Feng 2010). The sentiment classifier algorithms, using the given training data set, build a predictive model to classify new incoming data. In the absence of sufficient manually labelled data, a semi-supervised approach can be followed to extend the existing training data set with newly labelled instances.

Twitter Sentiment Analysis, using a machine learning approach, has been applied extensively. Some of the most applied sentiment classifiers include Naïve Bayes (NB), Support Vector Machines (SVM), Maximum Entropy (MaxEnt), Random Forests and Logistic Regression (da Silva & Hruschka 2014; Go et al. 2009; Pak & Paroubek 2010). An important drawback of supervised learning is that it tends to be domain-dependent and requires labelling of data in a new domain, or re-training new arriving data (Taboada 2016). On the contrary, once a labelled data set is available, that is, where text has been labelled positive, negative or neutral, training is trivial, and a classifier can be built relatively quickly with programming languages such as Python, which supports machine learning algorithms (Pedregosa et al. 2011; Perkins 2014; Sarkar 2016).

Unlike sentiment classifiers, the lexicon-based approach does not require training data, but instead relies on a sentiment lexicon. The lexicon-based approach is a rule-based approach and is used to analyse text at the document or sentence level in conventional texts such as blogs, forums and product reviews (Ding, Liu & Yu 2008; Kim & Hovy 2004; Turney 2002). The lexicon-based approach can be used across different domains without changing the dictionaries, making it an attractive approach for TSA (Taboada et al. 2011). In this approach, sentiment values of text are derived from the sentiment orientation of the individual words using an existing lexicon dictionary. The sentiment values provided by the model's dictionary indicates a word's polarity (e.g. *awesome* is positive and *horrible* is negative). When new text is classified, words in the text are matched to words in the dictionary, and using various algorithms, the values are aggregated into a sentiment score for the text. In general, lexicon-based approaches are more intuitive, robust and easier to implement than supervised learning approaches. For example, lexicon-based approaches have been shown to be successful on conventional text, as well as tweets (Mohammad, Kiritchenko & Zhu 2013; Thelwall, Buckley & Paltoglou 2012; Thelwall et al. 2010). However, unlike the machine learning approach, lexicon-based methods are less explored in TSA, TSA is less explored mainly because of the uniqueness of tweet messages (words such as *gr8* and *yolo*) and the dynamic nature with new hashtags emerging daily (Giachanou & Crestani 2016).

Related work

Sentiment analysis is often used in recommendation systems, online advertising systems and question-answering systems (Pang & Lee 2008). More recent studies include business and governments mining opinions from human-authored documents to assist with reputation management (Seebach, Beck & Denisova 2012). Other applications include mining Twitter data for opinions and sentiments during political elections (Tumasjan et al. 2010), stock market indicators (Bollen, Mao & Zeng 2011) and identifying social issues during natural disasters (Neppalli et al. 2017). In addition to these, sentiment analysis can also explore how news events affect public opinion. For example, in a study conducted by Wang et al. (2012), a real-time sentiment analysis model was used to evaluate responses to and public opinion regarding the 2012 US presidential election, while a more recent study by Jiang, Lin and Qiang (2016) assessed public opinion during the whole life cycle of a large hydro project. In addition to these studies, real-world applications, such as We Feel, are made available to the public to gauge and explore the real-time signal of the world's emotional state (Milne et al. 2015). However, in South Africa the sentiment analysis of microblogging data has received very limited attention (see Ridge, Johnston & O'Donovan 2015; Swart, Hardenberg & Linley 2012).

Methodology

Corpus

Twitter provides two application programming interfaces (APIs) to access and collect data – REST API and Streaming API. The Twitter Search API is part of Twitter's REST API and allows search against a sample of recent tweets published in the past 7 days. The Twitter Streaming API, on the contrary, provides developers' low-latency access to Twitter's global stream of Tweet Data over a longer period. Both the Twitter Search API and Twitter Stream API were used to collect a sample data set. A Python 2.7 tweet collector application was developed for the Twitter Streaming API to collect the responses from the public stream, all in JSON format. Twitter Archiver, a publicly available plugin for Google Sheets, was used to collect and archive tweets from the Twitter Search API (Agarwal 2015). As the study focuses on the minority community of Orania, tweets were downloaded with the keyword *orania* or hashtag #*orania* in both APIs. Sample data were collected over a single month (09 September 2017 to 09 October 2017) using both APIs. The Twitter Streaming API tweet application collected 272 tweets, while Twitter Archiver collected 895 tweets. As the study's aim is to gain a comprehensive understanding of the public sentiments about a minority community, the Twitter Archiver was used for further data collection. In total, 10 104 tweets were collected and archived between 09 September 2017 and 15 March 2018. The data set was then filtered according to language, where only English tweets were selected using Google's DetectLanguage() function. This function is offered as part of Google's Spreadsheet function list and calls the Google Translation API to detect the language of a string parameter

(Google 2018). Next, all tweets unrelated to the Orania community were removed. The final corpus consisted of 7192 tweets, with 5309 unique tweeters and an average number of words of 16.92 words per tweet.

Text preprocessing

Basic linguistic preprocessing is required to prepare the lexical source for sentiment analysis, as most forms of social media (except reviews) are very noisy. This function includes preprocessing tasks and methods that include data cleansing, tokenisation and syntactic parsing (Dey & Haque 2009). Firstly, external links and user names (signified by @ sign) were eliminated. We replaced all URLs with a tag ||HTTP_URL|| and targets (e.g. '@John') with tag ||AT_USER||. Special care was also taken with elongated words. We replaced a sequence of two or more repeated characters by two characters, for example we converted 'huuuuuungry' to 'huungry'. Special characters ($, % and #) and punctuation marks (full stops, commas, question marks and exclamation marks) were removed, except emoticons, as people often use the latter to express sentiment with tokens such as ':)', ':-)' or ':('. Because of the small corpus size, retweets (tweets that are re-distributed and start with 'RT') were kept in the corpus. We also applied automatic filtering to remove duplicate tweets, and tweets that were not written in English. After cleaning, we performed sentence segmentation, which separates a tweet into individual sentences. As is standard in NLP practices, the sentences were tokenised. Several tokenisers were investigated and the TweetTokenizer as part of the National Language Toolkit (NLTK) by Bird, Klein and Loper (2009) was found to be best suited for the study. The tokeniser handles emoticons, HTML tags, URLs, retweets, user mentions and Unicode characters correctly. Finally, all English stop-words (i.e. words that are common words with low discriminating power, e.g. *the*, *is* and *who*) were removed and the remaining tokens were converted to lowercase.

Negation and modifiers handling

Negation refers to the process of converting words from positive to negative, or negative to positive by using special words: *never*, *no*, *not* and *n't*. Handling negation is an important step of sentiment analysis as the negation can influence the sentiment of a text. A simple implementation strategy of handling negation was followed in this study: if a negation word is found, the polarity score of the word following the negation is reversed. A similar strategy was followed handling modifier words such as *very*, *much* and *really*. The polarity of the word following the modifier was adjusted with a factor of 1.3, thus increasing or decreasing the polarity of the word.

Sentiment analysis

The design of the sentiment model used in this study followed the lexicon-based approach using two lexicon dictionaries. This was because of the lack of sufficient training data and the notion that the lexicon-based approach can function without any corpus and does not require any

training. A Python 2.7 sentiment analysis classifier was developed to handle the preprocessing, negation, modifiers and score each sentiment. The sentiment score of each tweet was derived using the polarity scores of each word found in the lexicon dictionaries. The scores were then used in a classification method to classify the polarity of the tweets into either positive, negative or neutral categories.

Lexicons used

The sentiment analyser employed two lexicon-based dictionaries, namely Bing Liu's Opinion Lexicon (Hu & Liu 2004) and the National Research Council Canada (NRC) Hashtag Sentiment Lexicon (Mohammad, Kiritchenko & Zhu 2013). The Opinion Lexicon, assembled by Hu and Liu, consists of 6789 words and is divided into two lists of words: one containing positive ($n = 2006$) and one containing negative words ($n = 4783$). These words were manually extracted from customer reviews. Examples of positive opinion words are *beautiful, wonderful* and *good*, and examples of negative opinion words are *bad, poor* and *terrible*. The NRC Hashtag Sentiment Lexicon by Mohammad and colleagues consists of 54 129 unigram words associated with positive and negative sentiment and was generated automatically from tweets with sentiment-word hashtags such as #amazing and #terrible. The lexicon contains 32 048 positive and 22 081 negative words.

Score aggregation

Given a tweet *t*, the sentiment words were first identified by matching with the words in the two sentiment lexicons. We then compute an orientation score for the tweet *t*. Using, for example the lexicon of Hu and Liu (2004), a positive word is assigned the semantic orientation score of +1, and a negative word is assigned the semantic orientation score of −1. A similar approach is followed using the lexicon of Mohammad, Kiritchenko and Zhu (2013). The sentiment score of a tweet *t* is then calculated as the sum of scores of its sentiment words divided by the number of words with scores to produce an average score.

Evaluation measures

Precision, recall, *F*-measure and accuracy are evaluation metrics used to evaluate the performance of a sentiment analyser (Go et al. 2009). These evaluation metrics are used within a confusion matrix to indicate true positives (T_P), true negatives (T_N), false positives (F_P) and false negatives (F_N). True positives or negatives are correctly predicted values, which means the value of the tweet sentiment and the value of the predicted tweet sentiment are the same. False positives or negatives are incorrectly predicted values, which means the value of the tweet sentiment and the value of the predicted tweet sentiment are not the same.

Precision measures the ratio of correctly predicted positive instances among the identified positive or negative tweets. Precision (*P*) is defined as the number of true positives over the number of true positives plus the number of false positives (F_P). Recall measures the ratio of correctly predicted positive instances among all the positive or negative tweets. Recall (*R*) is defined as the number of true positives (T_P) over the number of true positives plus the number of false negatives (F_N). Accuracy is the most intuitive measure among these and measures the ratio of correctly predicted instances among all the tweets. Finally, the *F*-measure (or F1 score) is the weighted average of Precision and Recall. The formulas for the evaluation measures are given as follows:

$$\Pr ecision(p) = \frac{TP}{TP + FP} \qquad \text{[Eqn 1]}$$

$$Recall(r) = \frac{TP}{TP + FN} \qquad \text{[Eqn 2]}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad \text{[Eqn 3]}$$

$$F - measure = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \qquad \text{[Eqn 4]}$$

Method evaluation and results

To evaluate the accuracy, recall and precision of the polarity classification method, a publicly available data set available within NLTK was used. The Twitter Samples data set (Bird, Klein & Loper 2009) contains 20 000 annotated tweets that were collected in July 2015 from the Twitter Streaming API. The tweets were grouped into positive and negative tweets using happy emoticons (e.g. ':)', ':-)', '=)' and ':D') and sad emoticons (e.g. ':(', ':-(', '=(' and ';('). The two groups were of similar size, each consisting of 10 000 tweets.

A comparison was made to evaluate the effectiveness of the polarity classification method with other sentiment analysers, which included Pattern (De Smedt & Daelemans 2012) and AFINN (Nielsen 2011). Both sentiment analysers make provision for sentiment classification using a lexicon. The AFINN lexicon consists of 2477 words that have been purposefully created for sentiment analysis of microblogging messages such as tweets. The Pattern lexicon is a subjectivity lexicon-based on English adjectives, where adjectives have a polarity (negative or positive, −1.0 to +1.0) and a subjectivity (objective or subjective, +0.0 to +1.0) score. The results in Table 1 show that the lexicon-based classification method performed similarly when compared to other sentiment analysers.

TABLE 1: Summary of sentiment classification results (in percentages).

Method	Lexicon	Accuracy	Precision (Neg)	Precision (Pos)	Recall (Neg)	Recall (Pos)	*F*-measure (Neg)	*F*-measure (Pos)
AFINN	AFINN own built-in lexicon	63.54	61.89	65.73	70.50	56.58	65.91	60.81
Pattern	Pattern own built-in lexicon	63.65	60.93	68.17	76.08	51.22	67.67	58.49
Our method	NRC Hashtag	63.33	64.40	62.40	59.60	67.06	61.91	64.65
Our method	Opinion Lexicon	63.24	59.48	71.92	83.04	43.44	69.32	54.16

AFINN, An affective lexicon by Finn Årup Nielsen; NRC, National Research Council Canada; Neg, Negative; Pos, Positive.

On average, the polarity classification method using the NRC Hashtag Sentiment Lexicon scored 63.40% for precision, 63.33% for recall, 63.28% for *F*-measure with an accuracy of 63.33%. The results of the polarity classification method using the Opinion Lexicon were on average 65.70% for precision, 63.24% for recall, 61.74% for *F*-measure with an accuracy of 63.24%. These results were very similar to the results of the sentiment analysers being used in AFINN and Pattern. On average, AFINN scored 63.81% for precision, 63.54% for recall and 63.36% for *F*-measure, while Pattern on average scored 64.55% for precision, 63.65% for recall and 63.08% for *F*-measure. For the purpose of sentiment analysis, our method was used using the NRC Hashtag Sentiment Lexicon. The results will now be presented.

Sentiment analysis results

To gain an understanding into the data set, the corpus was first analysed in terms of word frequencies. Thereafter, a time-series analysis was conducted, followed by a sentiment analysis of the Twitter data.

Content analysis

A word frequency analysis revealed the top hashtags used in the corpus of English tweets. Some of the most popular hashtags included #languagepolicy (*n* = 112), #blackmonday (*n* = 56), #hoërskoolovervaal (*n* = 42), #ann7prime (*n* = 36), #dstv405 (*n* = 36), #eff (*n* = 35), #effmarch (*n* = 33), #afriforum (*n* = 19), #blackfriday (*n* = 12), #pagans (*n* = 11), #racists (*n* = 11) and #kalushi (*n* = 11).

A word frequency analysis was also performed on the words (excluding hashtags) used in the corpus. Some of the most popular words included *go* (*n* = 351), *people* (*n* = 337), *apartheid* (*n* = 268), *white* (*n* = 259), *land* (*n* = 169), *bursting* (*n* = 167), *allowed* (*n* = 153), *government* (*n* = 152), *university* (*n* = 131), *move* (*n* = 130), *solidarity* (*n* = 113), *racist* (*n* = 112) and *racists* (*n* = 95). To gain a better understanding about the sequence of words, tweets features based on word *n*-grams (4-gram) were created to identify underlying themes (see Table 2).

The following tweets recorded the highest number of retweets:

- "Another whites only settlement in South Africa Kleinfontein sister Town to Orania" (*n* = 1348)
- "@EFFSouthAfrica please table a motion to disband Orania, declare #AfriForum as a right wing movement and ensure they feel the heat" (*n* = 176)
- "South Africa: Orania Schools Bursting at the Seams" (*n* = 117)
- "AfriForum and Solidarity must go open their university in Orania. #LanguagePolicy" (*n* = 111)
- "Welcome to Orania a Whites Only Settlement in South Africa" (*n* = 95)
- "The Orania land was bought in the 80s to prepare for the end of apartheid. The ANC knew about it then. Madiba visited Verwoerd's widow there. They have a museum honouring all Apartheid presidents except de Klerk. Statue of Verwoerd looks over the town. I wrote a paper on this" (*n* = 75)

TABLE 2: n-gram word sequences and themes.

4-gram word sequence	Frequency	Theme
('South', 'Africa', 'Kleinfontein', 'sister') ('Kleinfontein', 'sister', 'Town', 'to') ('sister', 'Town', 'to', 'Orania') ('Another', 'whites', 'only', 'settlement')	1349	Mention of similar town
('disband', 'Orania', 'declare', 'AfriForum') ('right', 'wing', 'movement', 'and') ('wing', 'movement', 'and', 'ensure') ('ensure', 'they', 'feel', heat)	177	Racism
('Orania', 'Schools', 'Bursting', 'at') ('Bursting', 'at', 'the', 'Seams')	170	Afrikaans as a medium of education
('AfriForum', 'and', 'Solidarity', 'must') ('Solidarity', 'must', 'go', 'open') ('university', 'in', 'Orania', 'LanguagePolicy')	111	Afrikaans as a medium of education
('Welcome', 'to', 'Orania', 'a') ('Orania', 'a', 'Whites', 'Only') ('Whites', 'Only', 'Settlement', 'in')	95	Racism
('Orania', 'land', 'was', 'bought') ('80s', 'to', 'prepare', 'for') ('the', 'end', 'of', 'apartheid') ('ANC', 'knew', 'about', 'it')	76	Racism
('kind', 'of', 'town', 'white') ('town', 'white', 'people', 'move') ('to', 'when', 'they', 'miss') ('when', 'they', 'miss', 'apartheid')	66	Racism
('That', 'Orania', 'is', 'protected') ('constitution', 'Whats', 'even', 'more')	51	Racism

ANC, African National Congress.

- "Paarl is the kind of town white people move to when they miss Apartheid but dont want the commitment Orania requires" (*n* = 66)
- "You know what's scary? That Orania is protected by the constitution. What's even more scary is that the kids who go to primary and high school in Orania can go to any university in THE COUNTRY! Multiracial and all! And we think racism is going to fall? Funny!" (*n* = 52)

Tweet time-series analysis

From a preliminary time-series analysis, several target dates showed an above average number of tweets (> 24.81), which would suggest that the hashtag #orania or keyword *orania* were tweeted frequently that day on Twitter. The content of these tweets referred to specific news events in South Africa during the data collection period, which are given in Table 3.

Tweet trending analysis results

The tweet corpus was analysed for retweet trends (i.e. tweets that are retweeted over a short period of time). Results can be seen in Figure 1.

The tweets that were retweeted the most are presented in Table 4.

Sentiment analysis results

Because the proposed method produced a consistent level of accuracy in comparison with other sentiment analysers, the English tweet corpus was analysed by the polarity classification method. In addition to this, AFINN and Pattern were also used as sentiment analysers. A threshold of zero was used to classify the tweets into positive, neutral or negative groupings. In other words, if the score was 0, which indicates no sentiment value, the tweet was classified as neutral. A score of +0 was considered positive, and a score

TABLE 3: Political or news events.

Date	Number of tweets	Political or new event
30-10-2017	122	Black Monday, the nationwide protest against farm attacks
06-11-2017	95	An old news report was shared
07-11-2017	87	Same as previous day
20-12-2017	81	A sign in Orania was photographed and tweeted. The sign reads: 'Attention all white journalists from Europe: Please leave your prejudice at the entrance'
29-12-2017	179	The Constitutional Court ruling that English can be the only language of instruction at the University of the Free State
17-01-2018	124	School reopens and the protests at Hoërskool Overvaal start. Protests concern the use of Afrikaans as a medium of education and its perceived use as a medium of exclusion
20-01-2018	128	Protests continue at Hoërskool Overvaal
21-01-2018	121	A picture of grade 1 learners at Orania is shared on social media platforms
22-01-2018	84	Tweets concern the same picture as previous
24-01-2018	226	Interest in Orania continues in the wake of Hoërskool Overvaal and the picture of learners at Orania
25-01-2018	113	The same as previous
26-01-2018	99	The same as previous
09-02-2018	188	The tweet with the highest number of retweets is shared
10-02-2018	383	Same as previous
11-02-2018	851	Same as previous
12-02-2018	87	Same as previous
15-02-2018	72	Election of Cyril Ramaphosa as new President
16-02-2018	233	Same as previous
17-02-2018	98	Same as previous
27-02-2018	71	Motion passed in Parliament to consider Land Expropriation Without Compensation
28-02-2018	76	Same as previous
01-03-2018	77	Same as previous
02-03-2018	329	Same as previous
07-03-2018	94	Same as previous
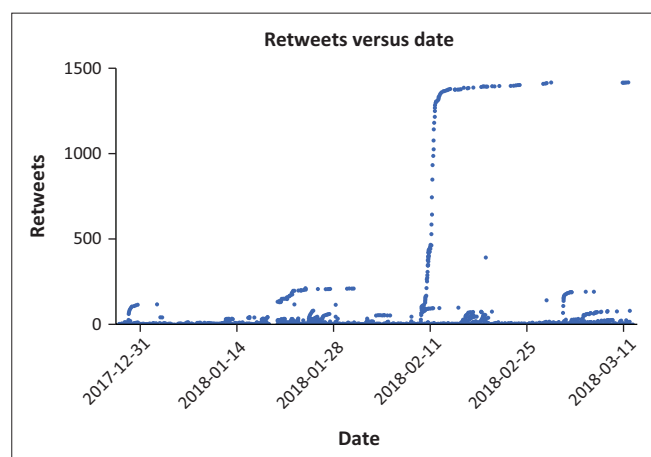
FIGURE 1: Retweet trends.

of −0 was considered negative. The results of the sentiment analyser in terms of polarities are shown in Tables 5 and 6, with some annotations in Table 7.

The varied results suggest that a lexicon is dependent on a particular domain. For example, the Lexicon Opinion words were extracted from customers reviews, which are not limited to 140 characters associated with Twitter messages. The Hashtag Sentiment Lexicon, on the contrary, was generated from tweets with sentiment-word hashtags, and thus, are much closer associated with the corpus used in

TABLE 4: Tweets that were retweeted the most.

Trending retweets	Trending from	Trending to	Retweet
RT @USER: "AfriForum and Solidarity must go open their university in Orania. #LanguagePolicy" https://t.co/0EWCwACEzc	29-12-2017	02-01-2018	114
RT @ USER: "South Africa: Orania Schools Bursting at the Seams"	19-01-2018	30-01-2018	208
RT @USER: "Another whites only settlement in South Africa Kleinfontein sister Town to Orania…"	09-02-2018	15-03-2018	1419
RT @USER: "@EFFSouthAfrica please table a motion to disband Orania, declare #AfriForum as a right wing movement and ensure they feel the heat!"	02-03-2018	06-03-2018	189
RT @USER: "The success of Orania carries many lessons for pro-European activists around the globe who seek the preservation and survival of their people"	04-03-2018	11-03-2018	77

EFF, Economic Freedom Fighters.

TABLE 5: Results of the sentiment analysis (with retweets, $n = 7192$).

Method	Lexicon	Negative	Neutral	Positive	Negative proportion (%)
AFINN	-	1324	2439	3429	18.41
Pattern	-	765	4777	1650	10.64
Own method	NRC Lexicon	3630	89	3473	50.47
Own method	Opinion Lexicon	1115	4632	1445	15.50

AFINN, An affective lexicon by Finn Årup Nielsen; NRC, National Research Council Canada.

TABLE 6: Results of the sentiment analysis (without retweets, $n = 2649$).

Method	Lexicon	Negative	Neutral	Positive	Negative proportion (%)
AFINN	-	685	1230	734	25.86
Pattern	-	427	1550	672	16.12
Own method	NRC Lexicon	1694	68	887	63.95
Own method	Opinion Lexicon	604	1542	503	22.80

AFINN, An affective lexicon by Finn Årup Nielsen; NRC, National Research Council Canada.

TABLE 7: Examples of sentiment classifications.

Twitter message	Sentiment classification
"I wish all racists people move to Orania"	Negative
RT @user: "The fact that the government has allowed Orania to exist is appalling"	Negative
@user "Orania is basically farm with around 1000 people living there. Hardly a threat. And there are less every year"	Neutral
"Orania offers a safe sanctuary from the crime-ridden neighbourhoods of South Africa, Orania is the first town of …"	Positive
@user "lies are spread about Afrikaners as a whole, nou just about the bunch in Orania. You're not alone on that one."	Positive

this study. We were surprised at the results of the AFINN, whose lexicon was also generated from tweets. The AFINN lexicon, however, only consists of 2477 words, while the NRC-Canada Hashtag Sentiment Lexicon consists of 54 129 unigram, and therefore, would be able to score more words than the AFINN lexicon. For these reasons, the study will use the results of the NRC lexicon in our discussion.

Discussion

The results above clearly show that Orania is associated with racism. Whether or not the community actually see themselves as a 'whites only' community, word frequencies and trending tweets clearly show an association with racism. The NRC lexicon's results, as shown in Tables 5 and 6 above, also indicate that the majority of unique tweets without retweets (63.95%), as well as total tweets (50.47%), have a negative sentiment towards Orania. Twitter is clearly used as a

platform to share negative sentiments about Orania. This is to be expected, as an Afrikaner-only community that goes against the dominant ideology and government policy of integration in a majority black country is bound to elicit fierce criticism. Note also that the word *racist* has a score of −1.377 in the NRC lexicon, while *apartheid* has a score of −4.999 and *racists* a score of −2.699; and given the high frequency with which these words occur in the corpus, a substantial part of the total negativity score can be attributed to the frequent occurrence of these negative words. Interestingly though, these negative tweets are all from outside the community: we checked for overtly racist tweets coming from users identified as people living in Orania and did not find a single occurrence. Racial slurs are limited to references to '*white pigs*', while no slurs targeting black people occur in this corpus.

The time-series analysis above also shows that Orania is part of the South African political landscape, especially where issues affect Afrikaners. The high number of tweets associated with the Black Monday protests on 30 October 2017, the issue surrounding the language policy at the University of the Free State, the issue around Hoërskool Overvaal and Afrikaans as a medium of instruction, the election of Cyril Ramaphosa as South Africa's new president and the debate around land expropriation are all issues that affect Afrikaners directly. Whenever a major event occurs that affects Afrikaners, mentions of Orania rise. In line with commentators such as Steward (2016) noting that anti-white sentiment on social media platforms is on the increase, negative sentiment towards Orania is also tied to negative sentiments towards Afrikaners. Black Monday is a case in point: the nationwide protest against farm murders was condemned as racist by the ANC, EFF and Black First Land First (BLF) (Khoza 2017a; Mphahlele 2017), and this was accompanied by tweets such as:

'Orania Racists are out in our streets today #BlackMonday'.

'Orania is basically an old apartheid flag, been provoking #BlackMonday'.

And:

'that @Username [from Afriforum] is just a piece of racist crap. He's representing a bunch of rightwingers who are bitter, they can go to hell or Orania if they so wish.'

However, our classifier had limited success when identifying the most negative tweets. The most negative tweets (-0.9 to−0.999) in this corpus, together with the polarity score of each word, are shown in Table 8. The words that contributed most to a tweet's negative polarity are in bold.

None of these tweets are exceedingly negative, with some actually positive. Tweet no. 10, for instance conveys a positive idea. Table 9 shows more examples of positive tweets that received a negative classification.

On the contrary, some exceedingly negative tweets were underestimated, as shown in Table 10.

Through our analysis, it became clear that more research needs to be performed on identifying hate speech and racist rhetoric.

TABLE 8: The most negative tweets.

Tweet text	NRC hashtag sentiment lexicon score	Polarity words
1. "I wish all racists people move to Orania"	-0.999	wish (-0.236) **racists (-2.699)** **people (-0.984)**
2. "The priority number 1 piece of land to be expropriated without compensation MUST BE Orania …"	-0.992	priority (-0.745) piece (-0.411) without (-0.434) **compensation (-4.999)** must (-0.292)
3. "Yey wena apartheid worshiper, his"	-0.981	yey (1.577) wena (0.478) **apartheid (-4.999)**
4. "Do have any idea how small orania is? Their currency or bills are backed by the rand. You can back"	-0.981	idea (-0.392) currency (-0.533) bills (-0.838) backed (-0.317) **rand (-4.999)**
5. "Victim Card Declined!! If you expect the same selective empathy you got 30yrs ago you are in"	-0.979	victim (-1.302) card (-0.258) declined (-2.16) expect (-0.672) selective (-1.467) **empathy (-1.718)** ago (-0.435)
6. "When is Fikile Mbalula going to cleanse Orania of crimes against humanity? Racism that is."	-0.972	crimes (-1.641) humanity (-1.448) **racism (-1.836)**
7. "Patriotism to the country always to the government only when it deserves it. If Cyril keeps this up i might just"	-0.964	**patriotism (-4.999)** government (-1.362) keeps (-0.189)
8. "Our Government has not even criminalised racial discrimination or banned the establishment"	-0.958	government (-1.362) not_even (0.683) **racial (-1.798)** discrimination (-1.467) banned (-1.589)
9. "You know what is scary? That Orania is protected by the constitution. What is even more scary is that the kids who go"	-0.952	know (-0.541) scary (-1.266) protected (-0.52) **constitution (-2.566)** even (-0.683) scary (-1.266) kids (-0.716)
10. "Whatever the reasons for Orania's self-reliance by Afrikaners, they have triumphed beyond ANC corruption of every"	-0.937	beyond (-0.322) anc (-0.214) **corruption (-4.999)**
11. "I don't understand why we still tolerate ORANIA. we tolerate racism?? who is protecting em? feeding em?US? WE SHOULD BE EMBARRASSED MZANSI????"	-0.923	understand (-1.258) tolerate (-1.186) racism (-1.836) protecting (-1.6) **embarrassed (-1.879)**
12. "Why is Orania allowed to exist?"	-0.921	allowed (-1.086) exist (-0.756)
13. "Orania and verbal diarrhoea??"	-0.918	verbal (-0.368) diarrhoea (-1.467)
14. "How is about we address the situation of Orania??? Or is that irrelevant??"	-0.914	address (-0.729) situation (-0.262) irrelevant (-1.75)
15. "Its look like the whole Orania is there. Is that Eugene de Kock??"	-0.912	like (-0.292) whole (0.109) **eugene (-4.999)**
16. "Twitter in orania !! but no auto-correct to help with spelling …"	-0.904	Twitter (-0.528) help (-0.646) **spelling (-1.539)**
17. "Such bravery these scum. Taking on helpless aged people. Every one of these cowardly cretins should be"	-0.904	bravery (-0.214) scum (-2.07) helpless (-2.033) people (-0.984) cowardly (-1.6) **cretins (-2.354)**

NRC, National Research Council Canada.
Note: The tweets used in the table were collected by the authors. Terms in bold are those with the highest impact.

Numerous tweets about Orania go beyond the sharing of a negative opinion and can be considered examples of abusive language. Future research could follow the line of research conducted by Tulkens et al. (2016) and work towards the automatic detection of abusive language, which is an especially important research area given the recent cases of hate speech and racism in the South African media. Note also that these negative

TABLE 9: Positive tweets with negative sentiments.

Tweet text	NRC hashtag sentiment lexicon score	Polarity words
1. "For all new followers we explain what is Orania, how it was established after the end of apartheid and what is the idea behind this unique town located in the Karoo."	-0.833	new (0.684) followers (-0.629) explain (-0.496) **apartheid (-4.999)**
2. "Orania collectively contributing more (e.g. taxes) to South Africa than they ever receive"	-0.769	collectively (0.255) **contributing (-4.999)** taxes (-1.952) south (0.387) africa (0.385)
3. "Many otherwise unemployable (in rest of South Africa) people earns a taxable living in Orania and similar areas. Once again benefitting SA fiscus"	-0.740	many (-0.595) otherwise (-0.638) rest (0.477) south (0.387) africa (0.385) people (-0.984) **earns (-4.999)**
4. "Orania does not perpetuate the killings of gays and lesbians"	-0.732	**killings (-1.13)** gays (-0.351) lesbians (-0.715)

NRC, National Research Council Canada.
Note: The tweets used in the table were collected by the authors. Terms in bold are those with the highest impact.

TABLE 10: Underestimated negative tweets.

Tweet text	NRC hashtag sentiment lexicon score
1. "How about we accidentally get our hands on some grenades and start genocide? Orania is a good place to start"	-0.644
2. "But it's ohk when you stupid settler morons from Holland attack a black president"	-0.626
3. "Shoot the boer! Down with the Nazi. Fuck the… "	-0.560
4. "The privilege continues, the pig skins never give up but we won't let go. EFF will stop these rubbish tutorials. If they can't be like others, they must go to Orania for now … before we take it as well"	-0.348
5. "Maybe in Orania and soon you will be out of the anthem FUTSEK POES. Hows thats for Afrikaans?"	-0.219
6. "That Orania shit place must be burnt down we want them back in Europe or Australia"	-0.071

EFF, Economic Freedom Fighters; NRC, National Research Council Canada.
Note: The tweets used in the table were collected by the authors.

tweets are *about* Orania: we did not find a single tweet of someone defending Orania using such racist or hateful language.

We should note a few important limitations of the study. Twitter is not representative of the general population: Twitter users tend to be young and urban, and hence, one cannot generalise our results to conclude that the general South African population regards Orania in a negative light. Furthermore, from manually examining the user profile pictures and usernames of the top 50 tweeters (people), we deduce that the vast majority of users are not Afrikaners. Hence, the negative sentiment towards Orania does not include the perspectives of a substantial number of Afrikaners themselves. A future study will investigate the views of Afrikaners, but that will involve using social media platforms other than Twitter, for example Facebook.

Conclusion

The rise of social media brought a wealth of data that can aid in the understanding of social issues. This article showed some of the potential and limitations when using sentiment analysis to extract meaning from unstructured text when trying to gauge the opinions people have of a community. In the case of Orania, it was shown how negatively this community is portrayed on Twitter. It was also shown how the discourse on this community is tied to the discourse on the

Afrikaner in general through the fact that mentions of Orania rise when major issues occur that concern the Afrikaner.

One of the most important avenues for future research that was identified in this study is the need to identify racist and hate speech. Some of the tweets clearly showed undertones of hate speech (see example 3 in Table 10). However, the automatic detection of abusive language online is an open challenge for NLP and still an emerging research field. Our study found that using lexicons as sentiment analysis technique was not sufficient in the automatic detection of hate speech, but rather the sentiment of the tweet. Possible future research could consider using collocation extractions, as most words are not offensive in themselves but become offensive with other words or word combinations. In particular, word embeddings (Mikolov, Yih & Zweig 2013) could be useful. Machine learning, which is a subfield of artificial intelligence, could also be used as a statistical approach to train an automatic detection system to 'learn by example'. Given the continuing and escalating racial divisions in South Africa, this avenue of research can open up new opportunities to gauge the level of tolerance or lack thereof that characterise the South African society.

Authors' contributions

E.K. was the project leader and was responsible for experimental and project design and performed most of the experiments. B.S. made a conceptual contribution and ensured the overall scientific rigour of the project.

References

Agarwal, A., 2015, *How to save tweets for any Twitter hashtag in a Google sheet*, viewed 12 March 2018, from https://www.labnol.org/internet/save-twitter-hashtag-tweets/6505/

Barbosa, L. & Feng, J., 2010, 'Robust sentiment detection on Twitter from biased and noisy data', *COLING '10 Proceedings of the 23rd international conference on computational linguistics: Posters*, August 23– 27, 2010, Beijing, China, pp. 36–44, Chinese Information Processing Society of China.

Bird, S., Klein, E. & Loper, E., 2009, *Natural language processing with Python,* O'Reilly, Sebastopol, CA.

Blenck, J. & Von der Ropp, K., 1977, 'Republic of South Africa: Is Partition a Solution?', *South African Journal of African Affairs* 7(1), 21–32.

Bollen, J., Mao, H. & Zeng, X., 2011, 'Twitter mood predicts the stock market', *Journal of Computational Science* 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Brink, E. & Mulder, C., 2017, *Rassisme, haatspraak en dubbele standaarde: Geensins 'n eenvoudige swart-en-wit-saak nie*, Solidariteit, Centurion.

Chen, H., Chiang, R.H.L. & Storey, V.C., 2012, 'Business intelligence and analytics: From big data to big impact', *MIS Quarterly* 36(4), 1165–1188.

da Silva, N.F.F. & Hruschka, E.R., 2014, 'Tweet sentiment analysis with classifier ensembles', *Decision Support Systems* 66, 170–179.

De Beer, F.C., 2006, 'Exercise in futility, or dawn of Afrikaner self-determination: An exploratory ethno-historical investigation of Orania', *Anthropology Southern Africa* 29(3), 105–114. https://doi.org/10.1080/23323256.2006.11499936

De Smedt, T. & Daelemans, W., 2012, 'Pattern for Python', *Journal of Machine Learning Research* 13(1), 2063–2067.

Dey, L. & Haque, S.M., 2009, 'Opinion mining from noisy text data', *International Journal on Document Analysis and Recognition (IJDAR)* 12(3), 205–226. https://doi.org/10.1007/s10032-009-0090-z

Ding, X., Liu, B. & Yu, P.S., 2008, 'A holistic lexicon-based approach to opinion mining', *Proceedings of the international conference on Web search and web data mining – WSDM '08,* February 11–12, 2008, ACM Press, Palo Alto, CA.

Eloff, T., 2017, 'Who owns the land?', 02 May, viewed 10 April 2018, from http://www.politicsweb.co.za/opinion/who-owns-the-land

Findlay, K., 2015, 'The birth of a movement: #FeesMustFall on Twitter', 30 October, viewed 18 February 2018, from https://www.dailymaverick.co.za/article/2015-10-30-the-birth-of-a-movement-feesmustfall-on-twitter/#.Wn1UvpP1Vn4

Geldenhuys, D., 1981, *South Africa's black homelands: Past objectives, present realities and future developments,* The South African Institute of International Affairs, Braamfontein.

Giachanou, A. & Crestani, F., 2016, 'Like it or not: A survey of twitter sentiment analysis methods', *ACM Computing Surveys* 49(2), 1–41. https://doi.org/10.1145/2938640

Go, A., Bhayani, R. & Huang, L., 2009, 'Twitter sentiment classification using distant supervision', *Processing* 150(12), 1–6.

Google, 2018, 'Detecting languages', 14 February, viewed 12 March 2018, from https://cloud.google.com/translate/docs/detecting-language

Gundecha, P. & Liu, H., 2012, 'Mining social media: A brief introduction', *2012 Tutorials in Operations Research. INFORMS*, 1–17

Hagen, L., 2013, 'A place of our own. The anthropology of space and place in the Afrikaner Volkstaat of Orani', Unpublished MA dissertation, UNISA.

Hermann, D.J., 2006, 'Regstellende aksie, aliënasie en die nie-aangewese groep', Unpublished PhD Thesis at the University of the North West, Potchefstroom.

Hu, M. & Liu, B., 2004, 'Mining and summarizing customer reviews', *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining – KDD '04,* August 22–25, 2004, ACM Press, New York, pp. 168–177.

Jeffery, A., 2018, 'Race rhetoric undermining race relations in SA – IRR', 20 March, viewed 09 April 2018, from http://www.politicsweb.co.za/documents/race-rhetoric-undermining-race-relations-in-sa--ir

Jiang, H., Lin, P. & Qiang, M., 2016, 'Public-opinion sentiment analysis for large hydro projects', *Journal of Construction Engineering and Management* 142(2), 1–12. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001039

Khan, J., 2014, 'The tribe living in isolation in orania', *The New Age*, 09 January, p. 10.

Khoza, A., 2017a, 'All lives matter, not just whites – ANC on #BlackMonday march', 30 October, viewed 12 March 2018, from https://www.news24.com/SouthAfrica/News/all-lives-matter-not-just-whites-anc-on-blackmonday-march-20171030

Khoza, A., 2017b, 'New social media research finds xenophobia rife among South Africans', 04 April, viewed 18 February 2018, from https://www.news24.com/SouthAfrica/News/new-social-media-research-finds-xenophobia-rife-among-south-africans-20170404

Kim, S.M. & Hovy, E., 2004, 'Determining the sentiment of opinions', *Proceedings of the 20th international conference on Computational Linguistics – COLING '04,* August 23–27, 2004, Association for Computational Linguistics, Morristown, NJ, p. 1367.

Kotze, N., 2003, 'Changing economic bases: Orania as a case study of small-town development in South Africa', *Acta Academica Supplementum* 1, 159–172.

Labuschagne, P., 2008, 'Uti Possidetis? versus self-determination: Orania and an independent "volkstaat"', *Journal for Contemporary History* 33(2), 78–92.

Lambsdorff, O.G., 1986, 'Teilung Südafrikas als Ausweg', *Quick*, 31 July, pp. 32.

McNally, P., 2010, 'Orania tourism: Come gawk at the racists', 01 February, viewed 20 September 2017, from http://thoughtleader.co.za/paulmcnally/2010/02/01/orania-tourism-come-gawk-at-the-racists/

Mikolov, T., Sutskeve, I., Chen, K., Corrado, G. & Dean, J., 2013, 'Distributed representations of words and phrases and their compositionality', in C.J.C. Burges, L. Bottou & M. Welling (eds.), *Advances in neural information processing systems 26: 27th Annual conference on neural information processing systems 2013,* pp. 3111–3119, Curran Associates, Inc., Lake Tahoe, NV.

Milne, D., Paris, C., Christensen, H., Batterham, P. & O'Dea, B., 2015, 'We feel : Taking the emotional pulse of the world', *Proceedings of the 19th Triennial Congress of the International Ergonomics Association,* 09–14th August, Melbourne.

Mohammad, S., Kiritchenko, S. & Zhu, X., 2013, 'NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets', *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013),* June 14–15, 2013, pp. 321–327, Association for Computational Linguistics, Atlanta, GA.

Mphahlele, M.J., 2017, '#BlackMonday: BLF slams 'racist' farm murder protest', 31 October, viewed 16 March 2018, from https://www.iol.co.za/news/politics/justice-safety/blackmonday-blf-slams-racist-farm-murder-protest-11785002

Muller, J.Z., 2008, 'Us and them. The enduring power of ethnic nationalism', 02 March, viewed 03 July 2017, from https://www.foreignaffairs.com/articles/europe/2008-03-02/us-and-them

Neppalli, V.K., Caragea, C., Squicciarini, A., Tapia, A. & Stehle, S., 2017, 'Sentiment analysis during Hurricane Sandy in emergency response', *International Journal of Disaster Risk Reduction* 21, 213–222.

Ngugi, F., 2017, 'Whites-only town in SA is a sign of continued white supremacy', 04 January, viewed 20 September 2017, from https://face2faceafrica.com/article/whites-town-sa-sign-continued-white-supremacy

Nielsen, F.A., 2011, 'A new ANEW: Evaluation of a word list for sentiment analysis in microblogs', *Proceedings Ttiel: Proceedings of the ESWC 2011 workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings,* May 30, 2011, pp. 93–98, CEUR-WS.org, Heraklion, Crete, Greece.

Osborne, S., 2018, 'South Africa votes through motion to seize land from white farmers without compensation', 01 March, viewed 10 April 2018, from https://www.independent.co.uk/news/world/africa/south-africa-white-farms-land-seizure-anc-race-relations-a8234461.html

Pabst, M., 1996, 'Partition: Still an issue in South Africa?' *Aussenpolitik* 47(3), 300–310.

Pak, A. & Paroubek, P., 2010, 'Twitter as a Corpus for Sentiment Analysis and Opinion Mining', in N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis et al. (eds.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation,* May 17–23, 2010, pp. 1320–1326, European Language Resources Association (ELRA).

Pang, B. & Lee, L., 2008, 'Opinion mining and sentiment analysis', *Foundations and Trends in Information Retrieval* 2(1), 1–135. https://doi.org/10.1561/1500000011

Pang, B., Lee, L. & Vaithyanathan, S., 2002, 'Thumbs up?: Sentiment classification using machine learning techniques', *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* July 6–7, 2002, pp. 79–86, Association for Computational Linguistics, Stroudsburg, PA.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al., 2011, 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research* 12, 2825–2830.

Pelzer, A.N., 1966, *Verwoerd aan die Woord,* Afrikaanse Pers-Boekhandel, Johannesburg.

Perkins, J., 2014, *Python 3 Text Processing With NLTK 3 Cookbook,* Packt Publishing, Birmingham.

Pienaar, T., 2007, 'Die aanloop tot en stigting van Orania as groeipunt vir 'n Afrikaner-volkstaat', Ongepubliseerde MA-verhandeling, Universiteit van Stellenbosch.

Ridge, M., Johnston, K.A. & O'Donovan, B., 2015, 'The use of big data analytics in the retail industries in South Africa', *African Journal of Business Management* 9(19), 688–703. https://doi.org/10.5897/AJBM2015.7827

Roets, E., 2017, 'Anti-white racism in South Africa', 16 February, viewed 10 April 2018, from http://www.politicsweb.co.za/opinion/antiwhite-racism-in-south-africa

Sarkar, D., 2016, *Text analytics with Python : A practical real-world approach to gaining actionable insights from your data,* Apress.

Schönteich, M. & Boshoff, H., 2003, *'Volk' faith and fatherland: The security threat posed by the White Right,* Institute of Security Studies, Pretoria.

Seebach, C., Beck, R. & Denisova, O., 2012, 'Sensing social media for corporate reputation management: A business agility perspective', *ECIS 2012 Proceedings,* June 11–13, 2012, Association for Information Systems, Barcelona, Spain.

Steward, D., 2016, 'Anti-white racism has turned virulent – FW de Klerk Foundation', 15 January, viewed 10 April 2018, from http://www.politicsweb.co.za/politics/antiwhite-racism-has-turned-virulent--fw-de-klerk-

Steyn, J.J., 2004, 'The "bottom-up" approach to Local Economic Development (LED) in small towns: A South African case study of Orania and Philippolis', *Town and Regional Planning* 47, 55–63.

Stieglitz, S. & Dang-Xuan, L., 2013, 'Social media and political communication: A social media analytics framework', *Social Network Analysis and Mining* 3(4), 1277–1291. https://doi.org/10.1007/s13278-012-0079-3

Sulzberger, C.L., 1977, 'Eluding the Last Ditch', *New York Times*, 10 August, viewed 18 July 2017, from http://www.nytimes.com/1977/08/10/archives/eluding-the-last-ditch.html

Swart, K., Hardenberg, E. & Linley, M., 2012, 'A media analysis of the 2010 FIFA world cup: A case study of selected international media', *African Journal for Physical Health Education, Recreation and Dance* 2, 131–141.

Taboada, M., 2016, 'Sentiment analysis: An overview from linguistics', *Annual Review of Linguistics* 2(1), 325–347. https://doi.org/10.1146/annurev-linguistics-011415-040518

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M., 2011, 'Lexicon-based methods for sentiment analysis', *Computational Linguistics* 37(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Thelwall, M., Buckley, K. & Paltoglou, G., 2012, 'Sentiment strength detection for the social web', *Journal of the American Society for Information Science and Technology* 63(1), 163–173. https://doi.org/10.1002/asi.21662

Thelwall, M., Buckley, K., Paltoglou, G. & Cai, D., 2010, 'Sentiment strength detection in short informal text', *The American Society for Informational Science and Technology* 61(12), 2544–2558. https://doi.org/10.1002/asi.21416

Tiryakian, E.A., 1967, 'Sociological realism: Partition for South Africa?' *Social Forces* 46(2), 208–221.

Tulkens, S., Hilte, L., Lodewyckx, E., Verhoeven, B. & Daelemans, W., 2016, 'A dictionary-based approach to racism detection in Dutch Social Media', *First Workshop on Text Analytics for Cybersecurity and Online Safety (TA-COS 2016),* May 23, 2016, pp. 11–17, Association for Computational Linguistics, Portorož, Slovenia.

Tumasjan, A., Sprenger, T., Sandner, P. & Welpe, I., 2010, 'Predicting elections with Twitter: What 140 characters reveal about political sentiment', *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media,* May 23–26, 2010, pp. 178–185, The AAAI Press, Washington, D.C.

Turney, P.D., 2002, 'Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews', *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL '02),* July 07–12, 2002, pp. 417–424, Association for Computational Linguistics, Philadelphia, PA.

Von der Ropp, K., 1979, 'Is Territorial Partition a Strategy for Peaceful Change in South Africa', *International Affairs Bulletin* 3(1), 36–47.

Von der Ropp, K., 1981, 'De republiek Zuid-Afrika: Een oplossing door deling van de macht of door deling van het land?', *Internationale Spectator* 35(2), 114–119.

Von der Ropp, K. & Blenck, J., 1976, 'Republik Südafrika: Teilung oder Ausweg?' *Aussenpolitik* 27(3), 308–324

Wang, H., Can, D., Kazemzadeh, A., Bar, F. & Narayanan, S., 2012, 'A System for real-time Twitter sentiment analysis of 2012 U.S. Presidential Election Cycle', *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, July 10, 2012, pp. 115–120, The Association for Computer Linguistics, Jeju Island, Korea.

Wixom, B. & Watson, H., 2010, 'The BI-based organization', *International Journal of Business Intelligence Research* 1(1), 13–28. https://doi.org/10.4018/jbir.2010071702

World Wide Worx, 2016, 'South African social media landscape 2017 – Executive summary', viewed 19 November 2017, from http://www.worldwideworx.com/wp-content/uploads/2016/09/Social-Media-2017-Executive-Summary.pdf