

## Research Article

# Exploring Links between Online Activism and Real-World Events: A Case Study of the #FeesMustFall

Yaseen Khan,<sup>1</sup> Surendra Thakur,<sup>1</sup> Obiseye Obiyemi ,<sup>1,2</sup> and Emmanuel Adetiba<sup>3,4</sup>

<sup>1</sup>KZN e-Skills CoLab, Durban University of Technology, Durban, South Africa

<sup>2</sup>Department of Electrical and Electronic Engineering, Osun State University, Osogbo, Osun, Nigeria

<sup>3</sup>Department of Electrical & Information Engineering and Covenant Applied Informatics and Communication Africa Center of Excellence, Covenant University, Ota, Nigeria

<sup>4</sup>Institute for Systems Science, Durban University of Technology, Durban, South Africa

Correspondence should be addressed to Obiseye Obiyemi; [obiseyeobi@gmail.com](mailto:obiseyeobi@gmail.com)

Received 18 November 2021; Revised 12 January 2022; Accepted 11 February 2022; Published 7 March 2022

Academic Editor: Kashif Zia

Copyright © 2022 Yaseen Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The declaration of free education for the poor and working class by the South African government remains a key result of the two-year #FeesMustFall movement. The campaign was tainted by a few heinous incidents that sought to disrupt its objectives. This study examines the connections between online activism and real-world occurrences by undertaking a longitudinal sentiment analysis of textual conversations on the Twitter platform since no quantitative research has been documented in this regard, particularly on the opinions of social media users and their associated impacts. Between October 15, 2015, and April 10, 2017, 576,583 tweets with the hashtag campaign #FeesMustFall were collected and analyzed using the Valence Aware Dictionary Sentiment Reasoner (VADER). The change point analysis (CPA) method was used to detect various changes in the dataset, and the cumulative sum analysis (CUSUM) method was used to discover changes over time. Results revealed that this online activism sentiment reacted to and reflected real-life events. The sentiment expressed is triangulated with a perceived real-life negative event, which is the burning of the University of Johannesburg (UJ) hall and the library at the University of KwaZulu-Natal (UKZN), thus confirming how Internet activism influenced these real-world events during the #FeesMustFall campaign. The study makes a significant contribution because it is the first longitudinal examination of the #FeesMustFall campaign's sentiment distribution and variations.

## 1. Introduction

As of 2017, the #FeesMustFall campaign is Africa's largest and longest-running social media campaign [1]. It began in South Africa in 2015 when underprivileged students were unable to pay their university fees, as well as the accompanying expenditures on housing, books, travel, and meals. According to the protestors, higher education has consolidated a new kind of class apartheid [2]. Even though it is best known as a student-led, nonpartisan protest movement, the movement drew support from people from all walks of life (including the wealthy, the poor, business, academia, and civil society), with various political parties attempting to secure additional mileage from the campaign [3–8]. After

several failed attempts to ease tensions with alternative solutions, the #FeesMustFall campaign appeared to reach a climax in 2017, when the government announced “Free Education” in mid-December 2017 [9,10]. During the campaign, #FeesMustFall made extensive use of social media [11], with Twitter gaining particular attention due to its hashtag function [12], which was included in the campaign's motto “#FeesMustFall.”

For over 15 years, Twitter has served as a platform for public discourse, promoting public debate with a large daily increase in the number of participants across various groups across the world [13]. Twitter is a prominent and accessible communication medium that generates a huge volume of data, mostly in the form of tweets, usually related to

discussions by a prefix or hashtag keyword (#). For connecting and updating conversation threads amongst users, hashtags provide a many-to-many communication tool [14]. Emoticons, emojis, photos, audio, and video can all be found in Twitter tweets. This type of data is now referred to as big data, and it offers a wealth of study opportunities for both historical and contextual analysis. Sentiment analysis is one sort of contextual analysis.

Social media has widely been validated as a tool for marketing, debates, teaching and learning, communication, campaigning, and public awareness, among other things. During the Arab Spring, for example, #Arabspring became the most-documented hashtag, with people using it in social media posts to show their support for the uprisings [15]. Microblogging systems like Twitter enhanced Japanese communication and rescue operations after an earthquake and tsunami in 2011 [16,17]. The floods in Queensland in 2010 and the earthquakes in Christchurch in 2011 both illustrate how the open service on Twitter helps with search and rescue operations by using spatial metadata while managing reactions [16,18,19].

These events have shown the usefulness of Twitter for communications during a crisis. Nonetheless, the 2011 London riots recorded operations of criminals facilitated by social media and BlackBerry Messenger's secure encrypted communication function to covertly foment violence and plan looting [20]. In the aftermath of the 2017 Barcelona terror attack, Twitter users teamed up to upload photographs of "cute cats" to prevent activists from circulating images of the dead and mayhem [21]. These examples underscore the disparity between Twitter's use-abuse, exploit-leverage, and social need for measures to monitor opinion expressed on the medium.

Scientific, corporate, and now government groups are becoming increasingly interested in capturing public opinion on social events, political movements, commercial strategies, and marketing activities. Approximately 95% of the data are classified as unstructured [22,23] and are rarely examined [24,25]. Sentiment analysis examines people's feelings, attitudes, and opinions, as well as their emotions, about things like products, people, subjects, organizations, and services. The market and perceived customer reactions to commercial brands have been the focus of most sentiment analysis studies. For instance, in [26], there is a focus on sentiment analysis and brand attitude, while [27] presents a decentralized approach to sentiment analysis studies, by evaluating sentiment polarity based on a dataset from Twitter with deducible occurrences in real life.

A recent study [28] was conducted to evaluate the various identities, communities, and discourses on Twitter during student protests. Specifically, between the 2015 #RhodesMustFall and the #FeesMustFall protests, 1000 tweets were collected per #OpenStellenbosch, #UCTFeesMustFall, #RhodesMustFall, and #WitsFeesMustFall hashtags. AntConc, a corpus linguistics software, was used to run the data, and the results were analyzed using critical discourse analysis (CDA). The study identifies Twitter as a medium for information exchange and is notable for

documenting various events related to the protest in real time. It also revealed that Twitter's major role was to serve as a supplement to conventional involvement rather than as a sole tool and more capable of synthesizing ideas and allowing people to integrate their personal lives and experiences into the public domain.

A related study [29] assesses the role of social media in the facilitation of effective student online activism. 567,533 tweets were examined as extracted from the founding years of the students' movements between 2015 and 2016, using a mixed research approach and prioritizing trend lines over headlines. The results highlighted a methodological hurdle for South African social science scholars and researchers in general, implying that stronger collaboration is required to ensure the long-term gains of the microblogging database management life cycle within the tertiary education ecosystem.

The #RMF campaign, also known as the Rhodes Must Fall campaign, is another prominent South African youth activism on Twitter [30]. The student-led #RMF protest at the University of Cape Town called for the removal of the statue of British colonialist Cecil John Rhodes, alleging that it fostered an exclusive society and entrenched racism, particularly among black South African students. An assessment of the qualitative content of tweets was performed with network analysis using NodeXL. The results of the analysis demonstrate how social media discussions can set mainstream news agendas and, as such, should not be viewed as detached from more traditional media platforms. The paper also asserted that youths are becoming more reliant on social media, which fosters the creation of a new crop of individuals distinguished by more personalized forms of activism.

In another interesting study [31], the influence of automated software robots during the #FeesMustFall campaign was evaluated. 85% of the harvested 576 823 tweets remained after cleaning, and the tweet behavior of 90 783 unique individuals was studied in terms of tweet source, frequency, topic, and volume. The DeBot API detected four bots in the tweets, while two suspicious bot or cyborg accounts were identified by supplementary approaches based on traits. The results established the presence of cyborgs and Bots, both playing a significant role and being accountable for the amplification of #FeesMustFall hashtags on Twitter.

Although the use of social media was critical in the coordination of the #FeesMustFall campaign, there is no literature that provides a quantitative insight into the attitudes of social media users throughout the campaign and the resultant real-life events. This study fills that void by examining the links between online activism and real-life events using the #FeesMustFall campaign and conducting a sentiment analysis of textual exchanges on the Twitter social media network over a reference period. The campaign #FeesMustFall is peculiar because it was tainted by several events that made the campaign unpleasant and attempted to undermine its intentions. The goal of the research is to assess the #FeesMustFall movement by looking at the association between Twitter sentiment polarity and real-life events.

## 2. Research Method

Twitter, like other social media networks, contains a vast amount of textual and nontextual data that may be analyzed using a variety of methodologies. This study explored the robust textual data and considered four possible analytical techniques, including content analysis (CA), thematic analysis (TA), social network analysis (SNA), and sentiment analysis (SA). After a detailed evaluation of the features of CA, TA, SNA, and SA, it was observed that CA, TA, and SNA were not suitable for this study. SA, on the other hand, was considered because of its potential for assessing the mood of online interactions and providing a clear picture of the emotions underlying the texts by separating them into positive, neutral, and negative tweets.

SA is the process of extracting essential information from unstructured text and transforming it into valuable business intelligence. SA is often used to mine opinions, attitudes, points of view, and emotions from tweets, text, speech, and data sources using natural language processing (NLP) [32]. The Linguistic Inquiry and Word Count (LIWC), IBM Watson Language Understanding, Patterns, Sentiwords, and several *Python* implementations, such as the Valence Aware Dictionary Sentiment Reasoner (VADER) and TextBlob, are only some of the available SA tools [33]. SA could employ either the lexicon-based approach or machine learning (ML), with Figure 1 depicting the decision tree showing the methodologies and approaches for SA [34].

ML may be used to automatically categorize tweets based on a pretrained dataset [35]. However, it was not considered for this study and is thus identified for future work. A sentiment lexicon, on the other hand, is a collection of lexical features, for example, words classified as either positive (delight, awesome, love, and success) or negative (harm, ugliness, grief, woeful, or worst) [36]. Although the creation and validation of such lists of opinion-bearing attributes are one of the most robust approaches for building trustworthy sentiment lexicons, it is also the most time-consuming. As a result, most applied sentiment analysis studies depend significantly on manually produced lexicon dictionaries that are preexisting, such as LIWC, Hu-Liu04 Dictionary, VADER, and GI [36–38]. This study employs a lexicon-based approach using the VADER technique.

Although the campaign started precisely on the 15th of October 2015, the data observation period spans from the 21st of March 2015, that is, the date #FeesMustFall was first mentioned, to the 10th of April 2017. The data collation process was achieved with support from Podargos, a professional data service provider, which assisted with the retrieval of credible longitudinal data for the #FeesMustFall movement with the provision of five hundred and seventy-six thousand, five hundred and eighty-three (576,583) tweets or data points. This remained a viable option and was carefully explored since Twitter limits direct access to their data. The data collation was strictly based on the acquisition of posts containing the “#FeesMustFall” hashtag and not case sensitive. The dataset contained tweets along with their metadata, delimited in

the UTF-8 format and capturing emoticons as well as emojis, while data contents in video, picture, and audio formats were not harvested because VADER could not analyze them.

The VADER module was used along with *Python* for the derivation of the sentiments, while the tweets were categorized into the positive, negative, and neutral modes based on the sentiment polarity ratings by [36] and, respectively, classified as  $PS > 0.05$ ,  $PS < -0.05$ , and  $-0.05 \leq PS \leq 0.05$ , where  $PS$  is the polarity score. *Python* was employed for the analysis of the data, and specialized algorithms were built to extract the information into relevant datasets based on numerous search criteria. The choice of the VADER library for the extraction of sentiments is due to its appropriateness for microblog-like content and interactivity with *Python*.

Figure 2 shows sentiment assessment and the data classification techniques for this investigation. It started with the gathering of tweets into a dataset, which was preprocessed before being fed into the VADER model. Duplicated data were eliminated during the preprocessing phase, while the VADER model estimated and allocated equivalent sentiment polarity ratings to each tweet. This resulted in an output dataset containing unique entries along with additional fields to represent each tweet’s sentiment polarity scores and classification.

The change point analysis (CPA) on the outputted dataset is broken down into three distinct stages, and they are preprocessing, mining, and analysis, as shown in Figure 3. CPA was carried out using the bootstrapped cumulative sum analysis (CUSUM) technique, which is proven to be very useful and effective based on the application and performance from similar studies [40]. To identify a credible link between online sentiment and real-life occurrences, change point dates were triangulated with previous #FeesMustFall timelines that contain major and known actual events. Although the magnitude and computational connection of the link is outside the scope of this study, the triangulation results give plausible reasons and research prospects regarding the effect of social media sentiment on real-life occurrences and vice versa.

VADER has been identified and described as a simple rule-based algorithm for analyzing sentiment in social media content [36] and has been widely proven to be reliable for measuring and classifying sentiments expressed in tweets [41] and emails [42]. It evaluates a collection of lexical features to determine the text’s polarity and sentiment, for example, words that are labeled as positive or negative based on their semantic orientation. The algorithm behind VADER is therefore based on the concepts highlighted in Equations (1) to (3).

$$E \in [-4; 4], \quad (1)$$

where  $E$  represents the sentiment score per word. A word’s emotion intensity or sentiment score is assessed on a scale of -4 to +4, with -4 being the most negative and +4 representing the most positive. A neutral sentiment is represented by the midpoint 0 of the scale.

The following expression in Equation (2) is used to normalize the overall sentiment polarity score ( $PS$ ):

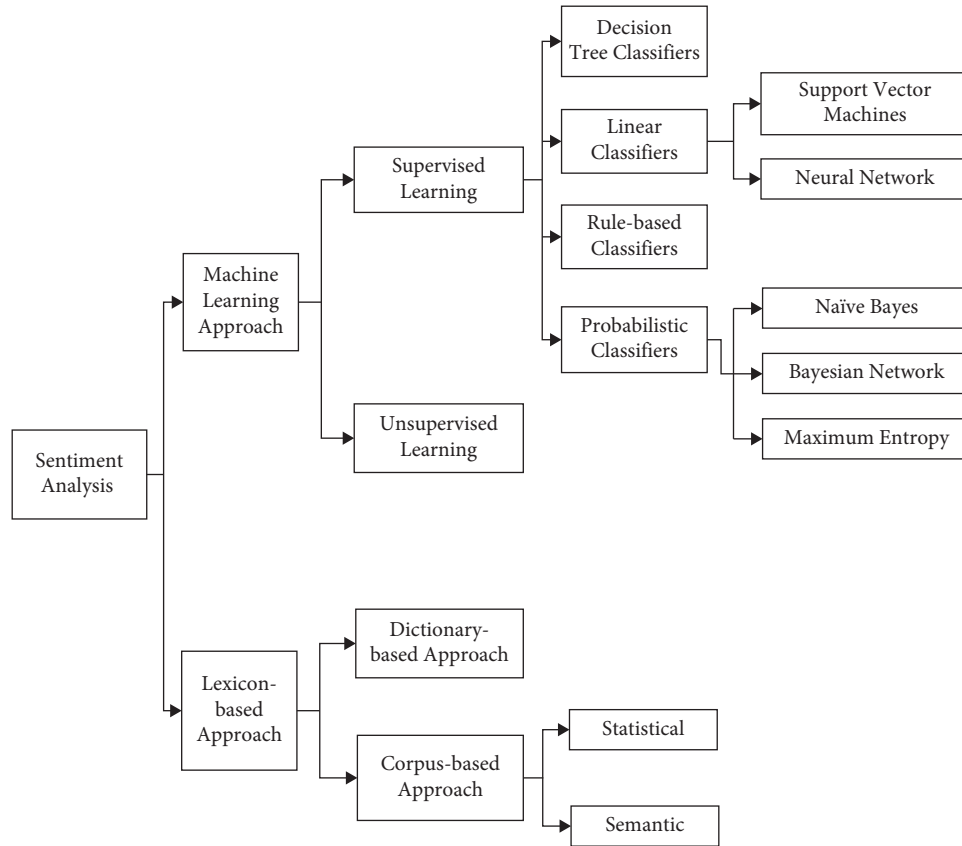


FIGURE 1: Techniques for categorizing sentiments (Medhat, Hassan, and Korashy, 2014).

$$PS = \frac{\sum E_i}{\sqrt{(\sum E_i)^2 + \alpha}} \quad i = 0, 1, 2, \dots, n, \quad (2)$$

$$PS \in [-1; 1], \quad (3)$$

where  $E_i$  represents the sentiment score of the  $i^{th}$  word, and  $PS$  is the overall sentiment polarity score. The normalization parameter  $\alpha$  is set to 15 by default.  $PS > 0.05$  is positive, and  $PS < 0.05$  is negative while  $-0.05 \leq PS \leq 0.05$  is neutral [36,43].

### 3. Results and Discussion

Table 1 shows the tweet distribution with a total of 490,449 tweets as collated over the observation period, whereas Table 2 shows the detailed statistics for the three categorized sentiments. The results show that positive sentiment accounted for 29.4% (144,026) of the overall dataset, with neutral sentiment accounting for 41.5% (203,693), while negative sentiment accounted for 29.1% (142,730) as shown in Table 2. Figure 4 shows the volume and percentage-based distribution of sentiment from tweets as generated from the monthly mean sentiments and the contents of Tables 1 and 2. As previously stated, March and April of 2015 were excluded since each had only one tweet. In addition, April 2017 was excluded because the period of study ended on April 10, 2017, leaving the month with somewhat inadequate days to estimate an acceptable average.

Figure 4 shows that the mineable data for October 2016 include a volume of around 80,000 tweets (Table 1 confirms that this is 82,712), with 26% positive tweets, 36% negative tweets, and roughly 38% neutral tweets. The broken vertical line shows two major perceived negative occurrences that happened during the #FeesMustFall movement, namely, the arson assaults on the UKZN library and the burning of the UJ Hall, respectively, recorded on day 6 and day 29 of September 2016. The leading tweeters and how they tweeted were appraised after additional analysis of the results. Table 3 shows the volume of the 10 most prolific tweeters along with other relevant parameters.

Table 3's columns are described as follows: The raw sentiment scores of each user's tweets were added and divided by the aggregate number of tweets to estimate the average sentiment score. The total number of hashtags (#) for a person is represented by No. of hashtags. No. of Favourite represents the sum of all user tweets which have been marked as favourites by a user, whereas No. of URLs in tweet indicates the overall number of URLs of a user. The number of a user's tweets is referred to as the "number of tweets." Retweets are the total number of user tweets that have been retweeted.

As indicated in Table 4, to ascertain the roles of news agencies, the analysis of news media usernames revealed the existence of five news media among the top ten tweeters. It demonstrates that the news media played a significant role in promoting #FeesMustFall tweets, with

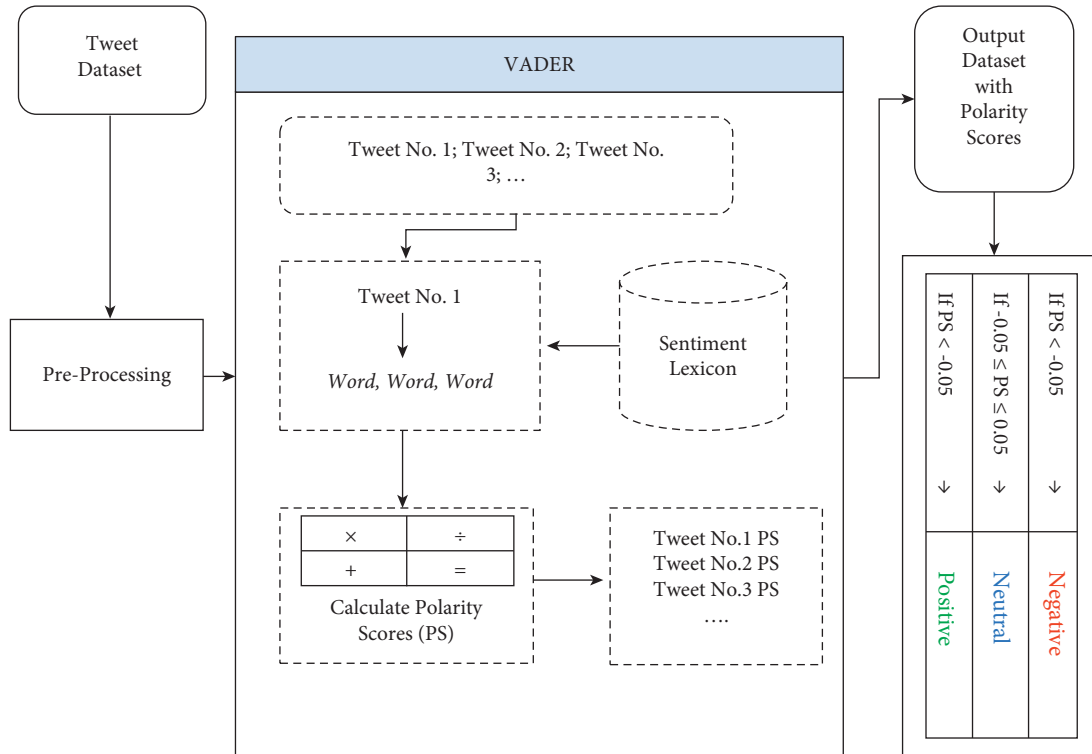


FIGURE 2: Sentiment analysis process for Twitter data [39].

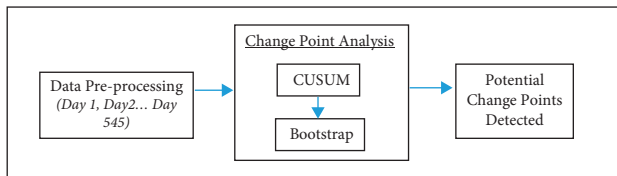


FIGURE 3: Change point analysis process flow [39].

2063, 1739, 1053, 1041, and 949 tweets attributed to Jacaranda News, EWN Reporter, The Daily VOX, POWER987News, and ANN7, respectively. Table 4’s columns are described as follows: The total number of tweets sent by a person is referred to as the “No. of Tweets”. The average number of hashtags per tweet is referred to as the “Hashtag Rate.” The average number of favourites per tweet is known as the “Fav Rate.” The average number of URLs per tweet is termed the “URL Rate.” The average number of retweets per tweet is known as the “RT Rate.” With a score of 2.73, ANN7 seems to be the only news media account with a hashtag rate higher than 1.94. Jou Ma Se Party has the lowest hashtag rate of any non-news media account, with a value of 0.24. Favourite ratings for news media accounts exceed 0.98. Non-news media accounts, on the other hand, had Favourite Rates of below 0.45. The Daily Vox is the only news media account with a URL rate of 1.04.

Altogether, the top five news media handles tweeted 6,485 times, with 58,986 RTs. With an RT rate of 16.66, EWN sent 1,739 tweets, which were retweeted 28,974 times. POWER987News and Daily Vox had RT rates of 6.36 and 11.43, respectively, highlighting the media’s role as

influencers. Interestingly, the top five non-news media records all had RT rates of less than 0.45, whereas the top five news media posts all had RT rates of more than 3.47.

**3.1. Tweeting Characteristics of the Highest Tweepers.** Camaren Peter was the most active tweeter, with 15,403 tweets that were favourited 488 times. There was an average of 4.14 hashtags per tweet, with one URL. As indicated in Table 3, this tweeter had the third lowest retweets with a total of 242 and recorded an average negative sentiment score of -0.07, which happens to be the second lowest. Figure 5 shows the monthly volume and average sentiment trend for Camaren Peter. The months of March and October 2016 were the busiest for this tweeter, with 1,915 and 3,829 tweets, respectively. An average sentiment of -0.13 and -0.14 was recorded, respectively, in the peak months of March and October 2016, while the most positive sentiment was recorded in March 2017 with an average score of 0.31. Another notable and high-ranking tweeter in this campaign is EduFunder, with 7,018 tweets that were favourited 146 times on his account. This is the smallest value in Table 4, indicating that these tweets had little influence. Table 3 shows that each tweet contained an average of 1.95 hashtags and was only retweeted 0.05 times, with a neutral sentiment score of 0.05.

**3.2. Overall Average Sentiment Polarity.** Mathematical tests and conditions stated earlier were used to analyze and evaluate the research objectives, which aim to identify how the online sentiment trends and polarity relate to the

TABLE 1: Distribution of tweets over a period of observation.

Month	Number of tweets
March, 2015	1
April, 2015	1
October, 2015	289,458
November, 2015	13,452
December, 2015	3,922
January, 2016	13,318
February, 2016	7,215
March, 2016	3,898
April, 2016	2,076
May, 2016	900
June, 2016	1,551
July, 2016	2,238
August, 2016	6,541
September, 2016	38,472
October, 2016	82,712
November, 2016	9,505
December, 2016	3,626
January, 2017	4,113
February, 2017	2,244
March, 2017	2,843
April, 2017	2,363
TOTAL	490,449

TABLE 2: Statistics of sentiment categorization.

Sentiment	Count	%
Positive	144,026	29.4
Neutral	203,693	41.5
Negative	142,730	29.1
TOTAL	490,449	100

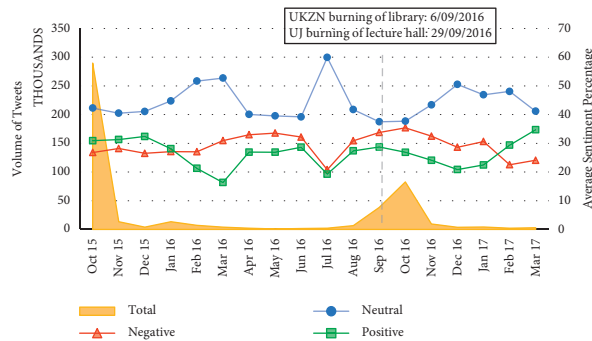


FIGURE 4: Sentiment analysis time series plot (by month) October 2015 to March 2017.

TABLE 3: The most prolific #FeesMustFall tweeters.

Username	Avg. Sentiment score	No. of hashtags (#)	No. of favourites	No. of URLs	No. of retweets	Number of tweets
Camaren Peter	-0.07	63,817	488	15,362	242	15,403
EduFunder	0.05	13,665	146	4,111	319	7,018
Wake up SA!!	0	15,684	1,013	2,215	1,025	2,318
Jou Ma Se Party	-0.09	533	56	2,294	70	2,258
#Africa	-0.09	7,355	426	2,221	100	2,193
Jacaranda news	-0.01	3,388	2,035	712	7,185	2,063
EWN reporter	-0.04	2,206	6,532	969	28,974	1,739
The daily VOX	-0.02	1,561	3,523	1,096	12,034	1,053
POWER987News	0	1,297	1,371	298	6,622	1,041
ANN7	-0.01	2,590	1,114	366	4,171	949

TABLE 4: Top 10 Tweeters and their respective rates.

Username	No. of tweets	Hashtag rate	Fav rate	URL rate	RT rate
Camaren Peter	15,403	4.14	0.03	1	0.02
EduFunder	7,018	1.95	0.02	0.59	0.05
Wake up SA!!	2,318	6.77	0.44	0.96	0.44
Jou Ma Se Party	2,258	0.24	0.02	1.02	0.03
#AFRICA	2,193	3.35	0.19	1.01	0.05
Jacaranda News	2,063	1.64	0.99	0.35	3.48
EWN Reporter	1,739	1.27	3.76	0.56	16.66
The Daily VOX	1,053	1.48	3.35	1.04	11.43
POWER987News	1,041	1.25	1.32	0.29	6.36
ANN7	949	2.73	1.17	0.39	4.4

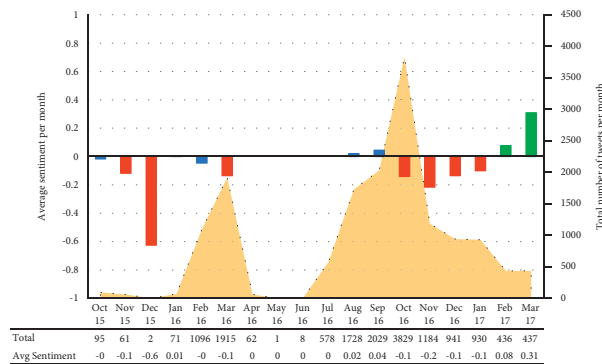


FIGURE 5: Camaren Peter's tweet distribution and average sentiment per month.

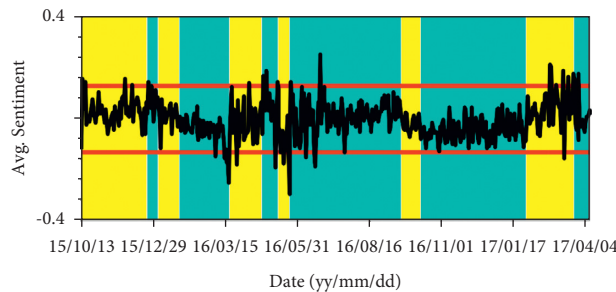


FIGURE 6: Plot of the average daily sentiment and CUSUM limits over the #FeesMustFall.

burning of the hall at the University of Johannesburg (UJ) and the library at the University of KwaZulu Natal (UKZN). As a consequence, 11 tweets were removed, leaving 490438 tweets to be analyzed throughout the 545-day observation period. The research results indicated there were variations in sentiment trend and polarity over the period of study. This is depicted graphically in Figure 6 and quantitatively in Table 5. Figure 6 presents a time series visualization of the combined daily average sentiment over the 545-day period. This is beneficial since it graphically displays changes with peaks and troughs, with average changes indicated in blue, while the maximum and lower CUSUM limits are represented by the red lines at the top and bottom. The graphic's backdrop color is yellow.

3.3. Identification of the Dates Signifying the Beginnings of Changes. The study was helpful in identifying the onset of sentiment trends and polarity shifts. Table 5 shows the

results of CPA using the CUSUM technique, which identified 11 change points. This finding indicates that there were variations in sentiment throughout the 545-day period of observation, with four (4) of the changes having a level of confidence of 100 percent. As shown in Table 5, the 11 date points when average sentiment polarity changed during #FeesMustFall indicate the change points with corresponding confidence levels. Real-life events that may be connected to these shifts were also found. Only two (2) dates, 20th March and 24th April 2016, were not connected to reliable real-life occurrences as witnessed, while the remaining nine (9) are connected with reliable incidences. The shift in sentiment was observed across the entire change points, with polarities ranging from negative to neutral to positive.

Apart from ascertaining if there were shifts in polarity and trend of negative sentiment during the observation period, it was also important to pinpoint the dates signifying the beginning of such changes, while isolating the change

TABLE 5: Changes in the average sentiment.

Date	Confidence interval (dd-mm-yy)	From	To	Confidence level (%)	Level
23-12-2015	(29-10-15, 31-12-15)	0.027693	0.077219	98	7
04-01-2016	(28-12-15, 11-01-16)	0.077219	0.012388	98	6
27-01-2016	(19-01-16, 11-03-16)	0.012388	-0.04478	100	9
20-03-2016	(28-01-16, 21-04-16)	-0.04478	-0.0052	99	5
24-04-2016	(25-03-16, 08-05-16)	-0.0052	0.060491	98	7
11-05-2016	(05-05-16, 15-05-16)	0.060491	-0.0997	100	8
23-05-2016	(15-05-16, 31-05-16)	-0.0997	0.014568	97	7
20-09-2016	(24-05-16, 29-09-16)	0.014568	-0.01755	96	8
10-10-2016	(26-09-16, 30-01-17)	-0.01755	-0.04232	100	3
01-02-2017	(28-01-17, 11-02-17)	-0.04232	0.052568	96	4
24-03-2017	(02-02-17, 03-04-17)	0.052568	0.004263	100	5

(Confidence level for candidate changes = 50%, confidence level = 95%, CI = 95%, bootstrap = 10 000, without replacement, MSE estimates).

TABLE 6: Significant changes for average negative sentiment.

Date	Confidence interval (dd/mm/YY)	Confidence level (%)	From	To	Level
07-08-2016	(30-07-16, 08-08-16)	99	-0.42085	-0.29499	7
06-09-2016	(05-09-16, 06-09-16)	100	-0.29499	-0.43423	6
08-09-2016	(08-09-16, 08-09-16)	98	-0.43423	-0.27649	9
19-09-2016	(19-09-16, 20-09-16)	100	-0.27649	-0.43862	5
10-10-2016	(24-09-16, 13-10-16)	97	-0.43862	-0.46454	7
01-11-2016	(31-10-16, 14-11-16)	99	-0.46454	-0.42244	8
22-03-2017	(09-03-17, 23-03-17)	95	-0.42244	-0.52847	7
29-03-2017	(27-03-17, 05-04-17)	98	-0.52847	-0.43271	8

(Confidence level for candidate changes = 50%, confidence level = 95%, CI = 95%, bootstrap = 10 000, without replacement, MSE estimates)

dates that related to real-life significant events. Interestingly, changes in polarity and the negative sentiment trend were found over the observation period. With the deployment of the CPA for the 142730 negative tweets, eight (8) significant change points were detected as shown in Table 6. The average negative sentiment (ANS) trend appears to have changed considerably over the 545 days.

Three (3) change points were correspondingly identified with relatable real-life events, and they were observed on the 6th of September, 19th of September, and 10th of October 2016. The 6th of September 2016 happens to be the same date when the library at UKZN was burnt, while the 19<sup>th</sup> of the same month was when Blade Nzimande announced the increase in student fees. Multiple real-life incidences occurred on the 10th of October 2016, with the demolishing of a bus at the University of Witwatersrand being the worst and most severe [44].

#### 4. Conclusions

In this research paper, the longitudinal sentiment analysis of the Twitter hashtag #FeesMustFall was conducted successfully. The study examined and analyzed the predominant sentiment of the opinions of Twitter users between 2015 and 2017 during the #FeesMustFall movement, while also quantifying the changes in sentiment in relation to prominent events recorded over the observation period, including the burning of the library at the University of KwaZulu Natal and the Hall at the University of Johannesburg, using computational techniques and methods to identify changes in sentiment trends.

The findings of the study revealed that the volume of prevailing sentiments may and does vary for several reasons, some of which are unknown, whereas others seem to react to or reflect events. The categorization of sentiment into negative, positive, and neutral groups offered additional contextual implications for the sentiment. Despite being among the top tweeters, the news media accounts displayed a surprising and largely neutral sentiment. To track changes in sentiment over time, the CPA analysis of both the average sentiment and the negative corpus of the tweets was utilized. The first CPA analysis at UKZN implemented CPA on all sentiment categories, but the second CPA analysis only employed the negative sentiment class, indicating the inherent advantage of stratified sentiment analysis. Studies reveal that changes in online sentiment patterns are closely connected to the unpalatable happenings both recorded at the University of Johannesburg and the University of KwaZulu Natal. On Twitter, CPA detected significant increases in negative sentiment trends, statistically reinforcing the notion that negative events have a compounding effect.

As shown in the findings of this study, sentiment analysis is an important process for understanding online activism like that of the #FeesMustFall campaign on Twitter. Unlike impractical and time-consuming manual procedures, the sentiment analysis techniques employed in this study are cost-effective and automated, providing realistic solutions for evaluating massive amounts of opinionated unstructured data commonly associated with the use of social media platforms. Furthermore, this study reveals how longitudinal sentiment analysis on Twitter may offer potential stakeholders important insights into historical trends in public



opinion. This is the first research of its type in South Africa to conduct a sentiment analysis study on the #FeesMustFall movement, giving information regarding the campaign's current sentiment patterns on Twitter and identifying the major change points particularly in relation to real-life events.

## Data Availability

The data that back up the outcomes of this study are readily accessible from the authors upon reasonable request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] D. Workman, *8 Hashtags that Shook Africa - IT News Africa - up to Date Technology News, IT News, Digital News, Telecom News, Mobile News, Gadgets News, Analysis and Reports*, <https://www.itnewsafrika.com/2017/02/8-hashtags-that-shook-africa/>, 2017.
- [2] J. A. Heher, "Report of commission of Inquiry into the feasibility of making high education and training fee-free in South Africa | the presidency," Available at: <http://www.thepresidency.gov.za/press-statements/release-report-commission-inquiry-feasibility-making-high-education-and-training>, 2017.
- [3] G. Makhafole, "Political parties support #FeesMustFall," 2015, Available at: <https://www.iol.co.za/news/politics/political-parties-support-feesmustfall-1932858>.
- [4] S. Mqadi, *EWN Ziyanda Ngcobo Compares #FeesMustFall 2015 and 2016 Dynamics*, Available at: <http://www.702.co.za/articles/235552/comparing-feesmustfall-2015-and-2016-dynamics>, 2015.
- [5] S. Booysen, *Fees Must Fall: Decolonisation, Higher Education and Governance in South Africa*, WITS University Press, 2016.
- [6] F. G. K. Buttelli and C. Le Bruyns, "# FeesMustFall as social movement and emancipatory politics? Moving towards an apocalyptic theological praxis outside the limits of party politics," *HTS Theological Studies*, vol. 73, no. 3, pp. 1–7, 2017.
- [7] M. Langa et al., et al. # Hashtag: An Analysis of the # Fees-MustFall Movement at South African Universities, 2017.
- [8] N. G. Wessels, *#FeesMustFall: Discourse Hidden in plain Sight*, 2017.
- [9] H. V. Mlambo, M. Hlongwa, and M. Mubecua, "The provision of free higher education in South Africa: a proper concept or a parable?" *Journal of Education and Vocational Research*, vol. 8, no. 4, pp. 51–61, 2017.
- [10] J. Zuma, *President Jacob Zuma: 2017 State of the Nation Address | South African Government*, 2017, Available at: <https://www.gov.za/speeches/president-jacob-zuma-2017-state-nation-address-9-feb-2017-0000>.
- [11] S. Jacobs and H. Wasserman, "The day mainstream media became old in South Africa - the Washington Post," 2015, Available at: <https://www.washingtonpost.com/news/monkey-cage/wp/2015/11/25/the-day-mainstream-media-became-old-in-south-africa/>.
- [12] S. Perez, "Twitter's doubling of character count from 140 to 280 had little impact on length of tweets | TechCrunch," 2018, Available at: <https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/>.
- [13] K. Bell, "15 years in, Twitter is ready to be more than just tweets | Engadget," 2021, Available at: <https://www.engadget.com/twitter-15th-anniversary-140008110.html>.
- [14] A. Bruns and J. Burgess, "ausvotes: how Twitter covered the 2010 Australian federal election," *Communication, Politics & Culture*, vol. 44, no. 2, pp. 37–56, 2011.
- [15] T. Rorke, D. Mendez, and L. Ceja, "#ArabSpring and mobilisation through social media - #MoveMe," Available at: <https://moveme.berkeley.edu/project/arab-spring/>, no date.
- [16] A. Acar and Y. Muraki, "Twitter for crisis communication: lessons learned from Japan's tsunami disaster," *International Journal of Web Based Communities*, vol. 7, no. 3, pp. 392–402, 2011.
- [17] S. E. Cho, K. Jung, and H. W. Park, "Social media use during Japan's 2011 earthquake: how twitter transforms the locus of crisis communication," *Media International Australia*, vol. 149, no. 1, pp. 28–40, 2013.
- [18] A. Bruns and Y. E. Liang, *Tools and Methods for Capturing Twitter Data during Natural Disasters*, First Monday, 2012.
- [19] A. Bruns and S. Stieglitz, *Metrics for understanding communication on Twitter*, Twitter and society [Digital Formations], vol. 89, , pp. 69–82, 2014.
- [20] R. Millham and S. Thakur, 'Social media and Big Data', 2016). *The Human Element of Big Data: Issues, Analytics, and Performance*, pp. 179–194, 2016.
- [21] A. Woods, "Twitter users flood Barcelona hashtags with cute animals to thwart terrorists," 2017, Available at: <https://nypost.com/2017/08/18/twitter-users-flood-barcelona-hashtags-with-cute-animals-to-thwart-terrorists/>.
- [22] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution that Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.
- [23] A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [24] J. Rizkallah, "The big (unstructured) data problem," 2017, Available at: <https://www.forbes.com/sites/forbestechcouncil/2017/06/05/the-big-unstructured-data-problem/?sh=1dc88775493a>.
- [25] C. Taylor, "Structured vs unstructured data 101: top guide | datamation," Available at: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>, 2021.
- [26] S. Kruijs and De Van, *Sentiment Analysis on Tweets. Measuring the Correctness of Sentiment Analysis on Brand Attitude*, Tilburg University, 2014.
- [27] P. Barnaghi, P. Ghaffari, and J. G. Breslin, "Opinion mining and sentiment polarity on twitter and correlation between events and sentiment," in *IEEE Second International Conference on Big Data Computing Service and Applications (Bigdataservice)*, pp. 52–57, IEEE, 2016.
- [28] A. Bolton, 'Tweeting in Solidarity': A Corpus Linguistics-Driven Analysis of Tweets during the South African Student Protests, University of the Witwatersrand, 2018.
- [29] N. Makhubu and A. Budree, "The effectiveness of twitter as a tertiary education stakeholder communication tool: a case of #FeesMustFall in South Africa," in *Social Computing and Social Media. Design, Human Behavior and Analytics Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 535–555, Springer-Verlag, 2019.

- [30] T. Bosch, "Twitter activism and youth in South Africa: the case of #RhodesMustFall," *Information, Communication & Society*, vol. 20, no. 2, pp. 221–232, 2017.
- [31] Y. Khan and S. Thakur, "The presence of twitter bots and cyborgs in the #FeesMustFall campaign," *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, IEEE, in *International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, pp. 1–5, 2018.
- [32] V. Kharde and P. Sonawane, *Sentiment Analysis of Twitter Data: A Survey of Techniques*, arXiv preprint arXiv:1601.06971, 2016.
- [33] O. Davydova, "Sentiment analysis tools overview, Part 1. Positive and negative words databases | by data monsters | medium," 2017, Available at: <https://medium.com/@datamonsters/sentiment-analysis-tools-overview-part-1-positive-and-negative-words-databases-ae35431a470c>.
- [34] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: a survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [35] M. Pennacchiotti and A.-M. Popescu, "A machine learning approach to twitter user classification," in *Fifth International AAAI Conference on Weblogs and Social media*, 2011.
- [36] C. Hutto and E. Gilbert, "Vader: a parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2014.
- [37] J. W. Pennebaker, M. E. Francis, and R. J. Booth, *Linguistic inquiry and word count: LIWC 2001*, Lawrence Erlbaum Associates, vol. 71, p. 2001, Mahway, 2001.
- [38] F. N. Ribeiro et al., et al. "Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods," *EPJ Data Science*, vol. 5, no. 1, pp. 1–29, 2016.
- [39] Y. Khan, *A Longitudinal Sentiment Analysis of the # Fees-MustFall Campaign on Twitter*, Durban University of Technology, 2019.
- [40] P. Monyeke, N. Naicker, and I. C. Obagbuwa, "Change-point Analysis: an effective technique for detecting abrupt change in the homicide trends in a democratic South Africa," *'Change-Point Analysis: An Effective Technique for Detecting Abrupt Change in the Homicide Trends in a Democratic South Africa', the Scientific World Journal The Scientific World Journal*, K. H. Ryu, vol. 2020, , pp. 1–10, 2020.
- [41] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and VADER sentiment," in *Proceedings of the International Multiconference of Engineers and Computer Scientists*, p. 16, 2019.
- [42] A. Borg and M. Boldt, "Using VADER sentiment and SVM for predicting customer response sentiment," *Expert Systems with Applications*, vol. 162, p. 113746, 2020.
- [43] P. Calderon, *Data Meets media*, Data Meets Media, 2017.
- [44] News24, "#FeesMustFall - shops looted in Braamfontein, clashes at UKZN - as it happened," 2016, Available at: <https://www.news24.com/news24/SouthAfrica/News/live-wits-fees-protest-turning-violent-20161010>.