We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Using Sentiment Analysis and Machine Learning Algorithms to Determine Citizens' Perceptions

Sherrene Bogle

Additional information is available at the end of the chapter

Abstract

This chapter analyzes the opinions expressed by individuals on four topical Jamaican issues and classifies them by emotions, feelings and polarity. The four trending topics on Twitter analyzed are the decriminalization of marijuana in Jamaica, Kaci Fennell's placing in Miss Universe, the Riverton Landfill fire and Barack Obama's working visit to Jamaica. The data pulled from Twitter for each topic was mined using three different classification algorithms to identify the accuracy of the data classified based on the polarity. The classifiers identified which polarity reflected what opinion is more dominant of the three; which are negative, positive or neutral. Sentiment analysis tools classified the opinions of Jamaican Twitter users with over 70% accuracy. Among three classification algorithms used, J48 decision tree received highest accuracy for the four topics tested and maintained the lowest error rate. For the decriminalization of marijuana, Kaci Fennell's placing in the Miss Universe competition and President Obama's visit, the accuracy was just over 70% and the mean absolute error (MAE) was less than 0.3. The methodology of the study provides a blueprint which can be utilized by managers and other decision making stakeholders to determine consumers' perception.

Keywords: Barack Obama, Jamaica, sentiment analysis, machine learning, Twitter

1. Introduction

Sentiment analysis uses linguistic and textual assessment, such as natural language processing to analyze word use, word order, and word combinations and thus to classify sentiments, often into the categories of positive, negative, or neutral polarity. Data gathered through sentiment analysis is believed to provide detailed information about something to which direct access did not previously exist: public opinion and feeling [1]. This research performs sentiment analysis by monitoring and analyzing local trending topics that create

a stir socially in Jamaica including President Barack Obama's historic working visit to the island country [2]. The aim of the study is to analyze the opinions and emotion expressed by citizens based on these topical issues and classifies them by emotions, feelings and polarity. It utilizes three machine learning algorithms to classify citizens perceptions namely decision tree J48, PART and naive bayes; and identifies the accuracy of the data classified based on the polarity. The classifiers identified the polarity reflected and which opinion is more dominant of the three (negative, positive or neutral). Research was undertaken on four topical issues in Jamaica: (1) The decriminalization of marijuana in Jamaica (2) Kaci Fennell's placing in the Miss Universe competition (3) The Riverton Landfill fire and (4) Barack Obama's working visit to Jamaica.

2. Methodology

The Sentiment analysis process consists of four main steps outlined in [3]: Data Acquisition, Data Pre-processing, Data Classification and Data Analysis.

2.1. Data acquisition

In this study, the twitter R package was used with RStudio to extract tweets which were subsequently used to create charts and classify data into emotions and polarity. Installation of packages such as install.packages ("twitteR", "ROAuth", "plyr") were required. The searchTwitter() function, found in the R library was used to obtain tweets on selected topics. Hashtags, single and double quotes were parameters accepted by the searchTwitter() function as a means of searching the Twitter API for tweets related to the keywords used in the search, for example temp = searchTwitter("#Jamaica Marijuana") would download tweets with the hashtag Jamaica Marijuana. It allows queries against the indices of recent or popular tweets and behaves similarly to, but not exactly like the search features available in Twitter mobile or web clients, making it very effective and easy to use in searching Twitter.

The population comprised of a corpus eleven thousand two hundred and five (11,205) tweets that were extracted from Twitter between January and April 2015. A search was done on Twitter to extract tweets on Jamaican topics that were not older than 2 weeks.

2.2. Data pre-processing

The corpus was also used offline where it was analyzed using machine learning and spreadsheet tools during pre-processing, classification and the post processing of the data. A function built into RStudio was then used to remove unwanted characters, texts, punctuations and numbers from the text files created as a result of the extracted data from Twitter. After successfully searching Twitter and obtaining the number of tweets required, the tweets were 'cleaned' using RStudio's cleaning function.

2.3. Data classification

RStudio provided two functions that analyzed the tweets and classified them into polarity (negative, neutral and positive) and emotion (joy, anger, fear, surprise). Analysis was done both on tweets (not re-tweeted) as well as re-tweets. After compiling the polarity function to classify the tweets into negative, positive and neutral polarities, the team observed that a number of tweets were classified incorrectly. This was a result of R's inability to understand the Jamaican dialect and RStudio's limited dictionary of words. Classifying tweets into emotions proved to be another challenge as majority of the tweets for the different topical issues returned a result of "unknown" for the emotion associated with the tweet. Both these tools, which are essential components of the sentiment analysis research being conducted, were somewhat ineffective in describing and classifying the data that was collected from Twitter.

2.4. Data analysis

The WEKA software was used offline to analyze data during pre- processing, classification and the post processing. In order to process the data gathered from the Twitter API, the file type or dataset was formatted to a file extension of .arff (attribute file format) and this file extension is generated from a.csv (comma-separated values) file which separates each attribute by a comma. The .arff file is an ASCII text file that describes a list of instances sharing a set of attributes.

A spreadsheet application was another useful tool in the sentiment analysis research conducted. This tool allowed one to inspect the comma-separated values files and also create graphs and tables.

3. Results

This section outlines the classification of tweets downloaded for each of the four topical issues. It shows how sentiment analysis of tweets can be used to explore the citizens perceptions on the topical issues selected. **Figure 1** below shows the summary and classification of such tweets.

Citizens from varying demographics express their opinions on Twitter on several topical issues. The four step methodology was executed and tweets were classified by the machine learning algorithms as shown in **Figure 1**. Barack Obama's visit to Jamaica represented the fourth bar among the quartet of tweets, received 2583 positive tweets and 658 negative tweets and the highest total tweets among the 4 topics investigated.

The next section will present results on polarity of topical issues for tweets and no-retweets.

3.1. Polarity of topical issues

A typical approach to sentiment analysis is to start with a lexicon of positive and negative words and phrases [4]. Polarity describes whether a word seems to evoke something positive
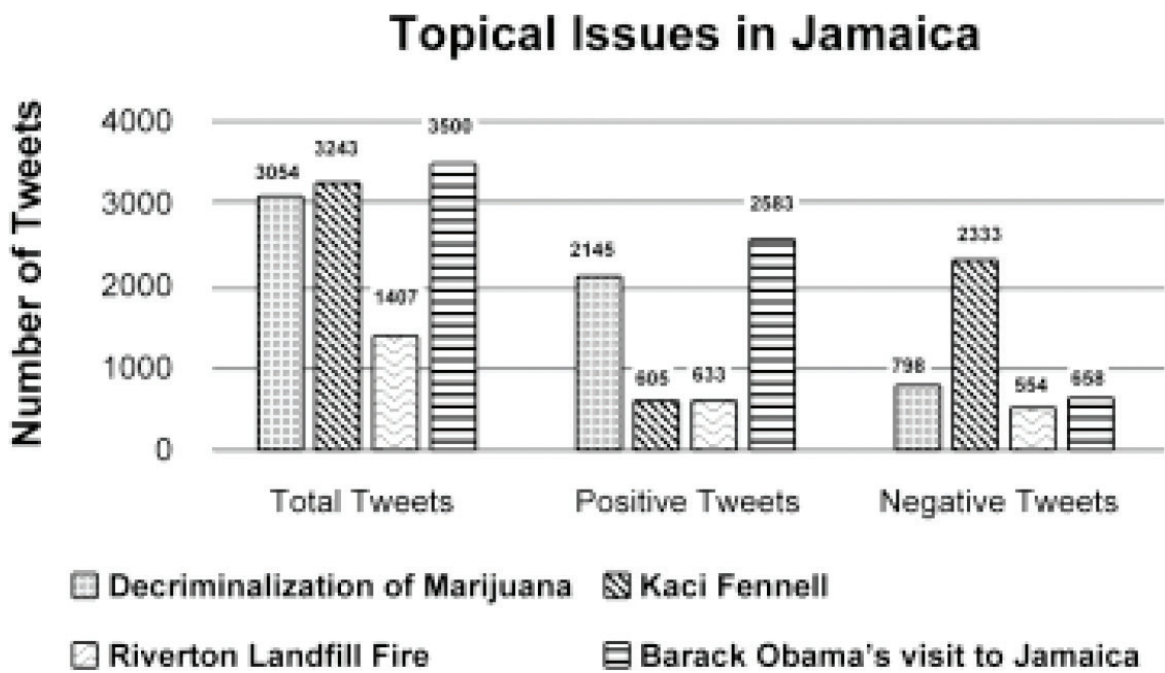
Topical Issues in Jamaica

Figure 1. Bar chart depicting the number of tweets collected on each topical issue.

or something negative. For example, *beautiful* has a positive polarity and *horrid* has a negative polarity. Examples of tweets that represent a positive polarity include:

Eg. 1: "Jamaica legalizes medical marijuana and decriminalizes recreational use."

Eg. 2: "the jamaican cabinet approves a bill to legalise use of small amounts of marijuana which will be examined in the senate this week."

Examples of tweets classified as having a negative polarity include:

Eg. 1: "what nbc didn't show kaci fennell miss jamaica"

Eg. 2: "miss jamaica says the miss universe pageant "went exactly as it should""

3.1.1. The decriminalization of marijuana in Jamaica

As depicted in **Figure 2**, majority (2145) tweets of the three thousand and fifty four (3054) tweets collected on the decriminalization of marijuana in Jamaica were positive. This demonstrates that the Jamaican citizens on the Twitter social media platform support the decision by government to decriminalize marijuana (*Cannabis sativa*) in Jamaica. However, seven hundred and ninety eight (798) Jamaicans on Twitter expressed negative sentiments toward the government's decision to decriminalize marijuana in Jamaica.

3.1.1.1. No-retweets

The graph above depicts the results obtained from analysis of tweets that were not retweeted, as in these tweets were posted by the original author. **Figure 3** shows that three hundred and forty six (346) tweets were negative and four hundred and seven (407) were positive.

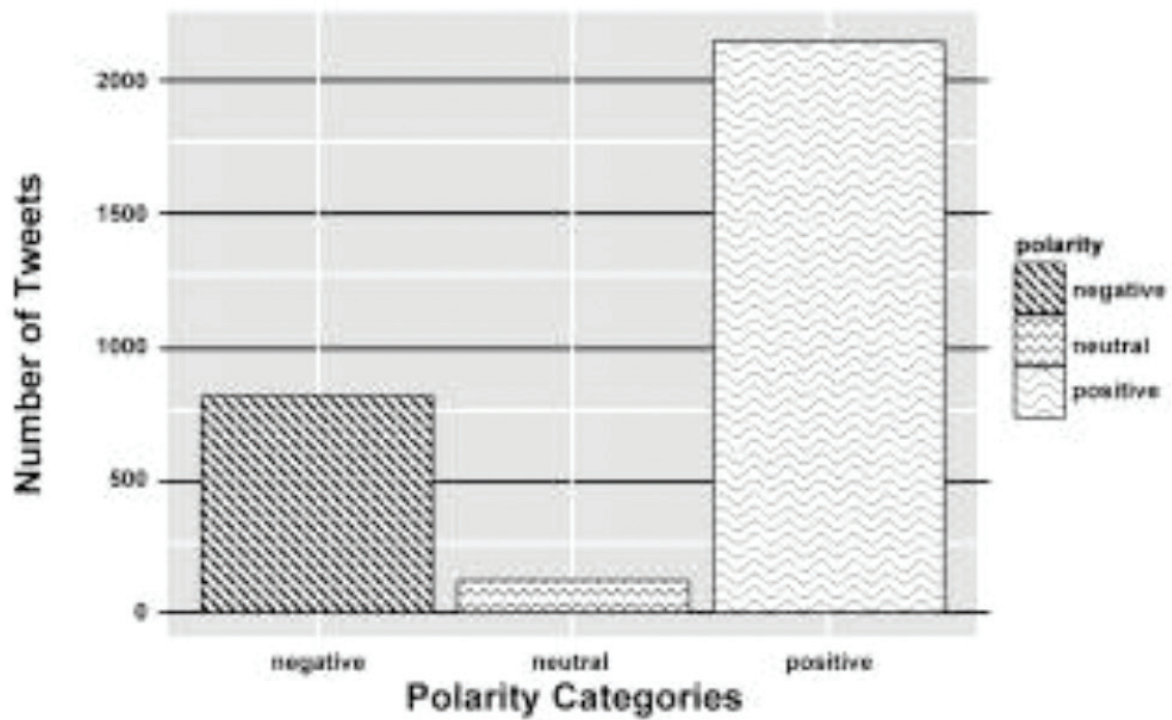Sentiment Analysis of Tweets on Decriminalization of Marijuana in Jamaica

Figure 2. Sentiment polarity of tweets obtained on the decriminalization of marijuana in Jamaica.

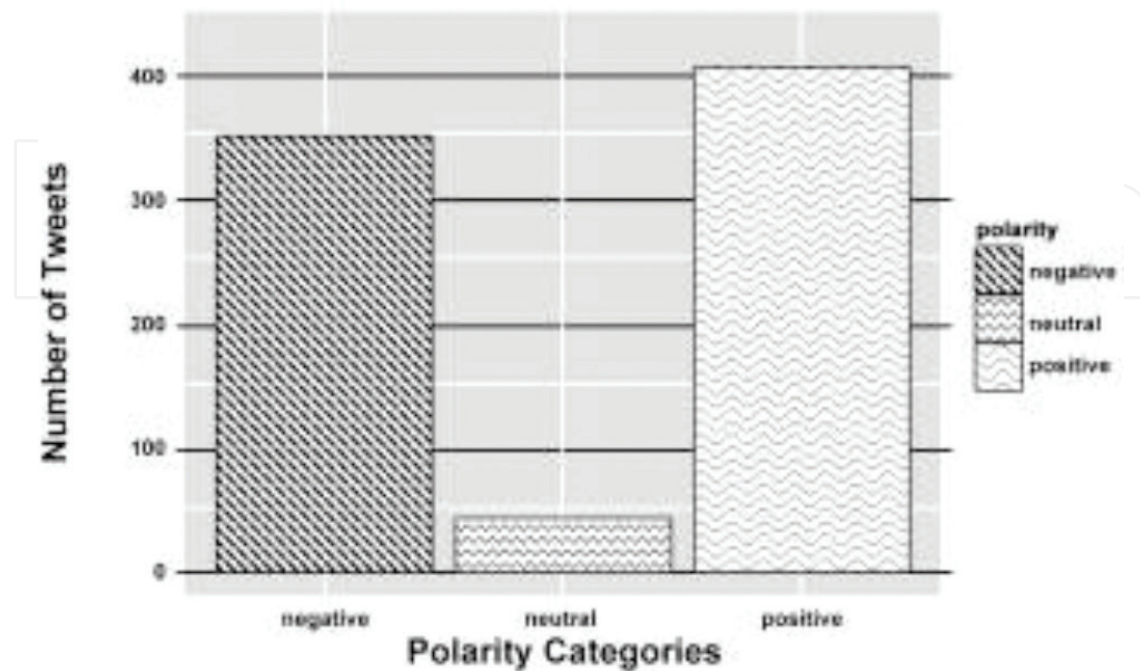Sentiment Analysis of No-Retweets on Jamaica Medical Marijuana

Figure 3. Sentiment polarity of no-retweets obtained on the decriminalization of marijuana in Jamaica.
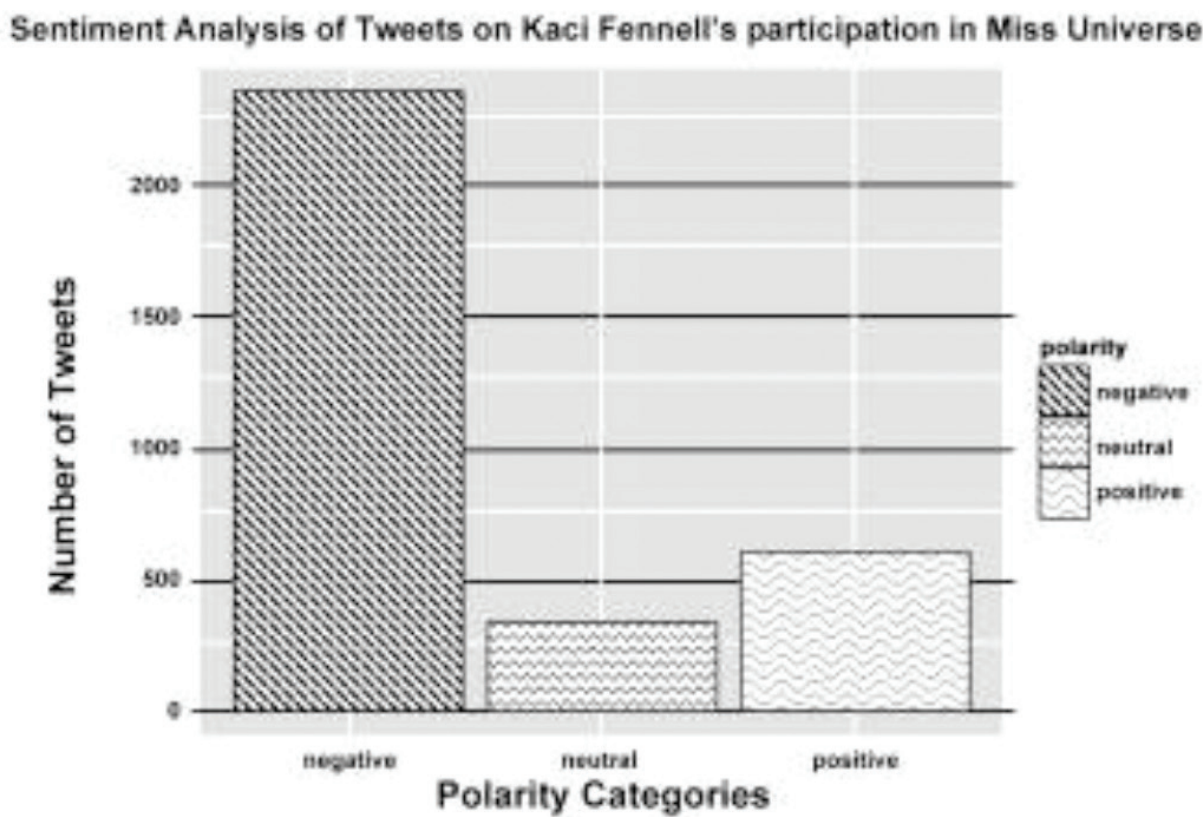
3.1.1.2. Retweets

Among retweets for the topic, there were four hundred and fifty three (453) tweets were negative and one thousand seven hundred and thirty eight (1438) were positive.

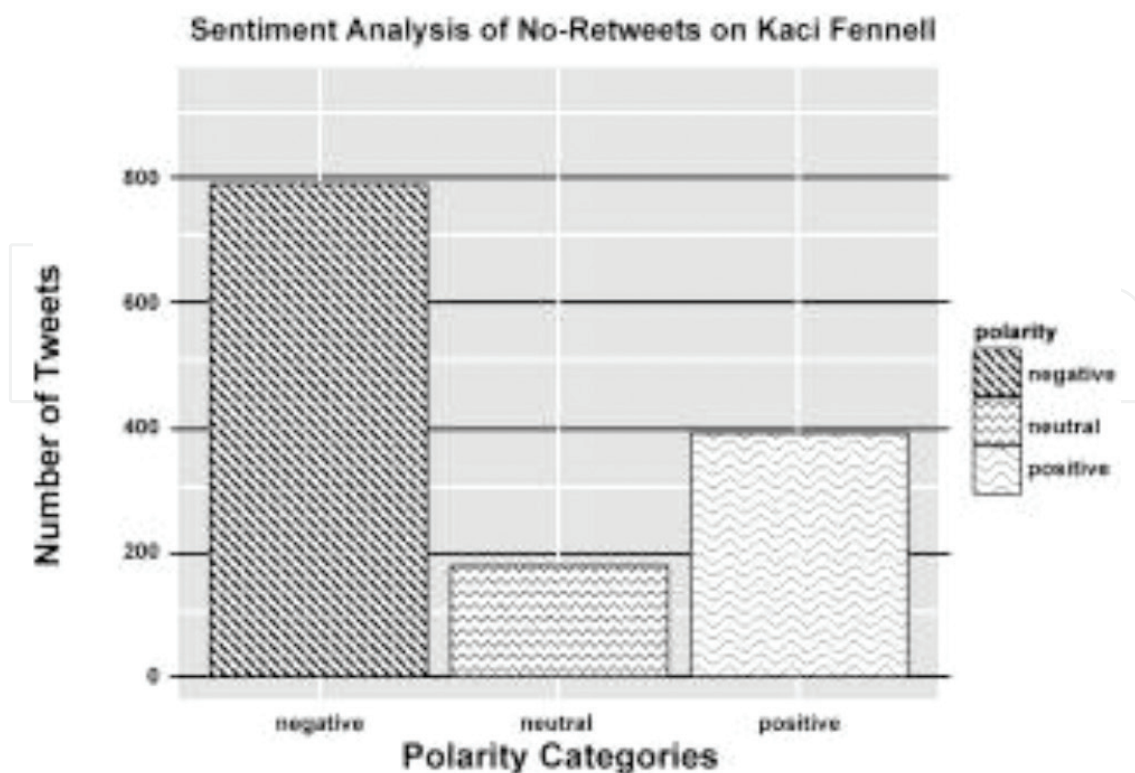*3.1.2. Kaci Fennell's placing in Miss Universe*

As depicted in **Figure 4** above, majority (2333) tweets of the three thousand two hundred and forty three (3243) tweets collected on Kaci Fennell's placing in the Miss Universe competition were negative. Upon examination of the tweets collected the negative tweets were expressions of anger and disappointment that Kaci did not win the Miss Universe competition or that she did not receive a higher placing than the fifth place ranking that she received. On the contrary, six hundred and five (605) tweets were classified as positive by the RStudio application.

3.1.2.1. No-retweets

The graph above depicts the results obtained from analysis of tweets that were not retweeted, as in these tweets were posted by the original author. **Figure 5** shows that seven hundred and ninety (790) tweets were negative and three hundred and ninety nine (399) were positive.

Figure 4. Sentiment polarity of tweets obtained on Kaci Fennell's placing in Miss Universe.

Figure 5. Sentiment polarity of no-retweets obtained on Kaci Fennell's placing in Miss Universe.

3.1.2.2. Retweets

Among retweets for the topic, one thousand five hundred and thirty (1530) tweets were negative and two hundred and forty (240) were positive.

3.1.3. Riverton Landfill fire

Figure 6 shows five hundred and fifty four (554) tweets of the one thousand four hundred and seven (1407) tweets collected on the Riverton Landfill fire in the Riverton community were negative. Smoke penetration from the fire was observed within a 20 mile radius from the landfill and further at times based on the wind direction. The fire lasted for 2 weeks and at least 29 critical air pollutants was detected [5]. The tweets classified as negative were Jamaicans expressing their anger toward the maintenance of the landfill and the effects of the fire on nearby communities. Six hundred and thirty three (633) tweets were classified as being positive.

3.1.3.1. No-retweets

The graph above depicts the results obtained from analysis of tweets that were not retweeted, as in these tweets were posted by the original author. **Figure 7** shows that two hundred and forty five (245) tweets were negative and three hundred and five (305) were positive.
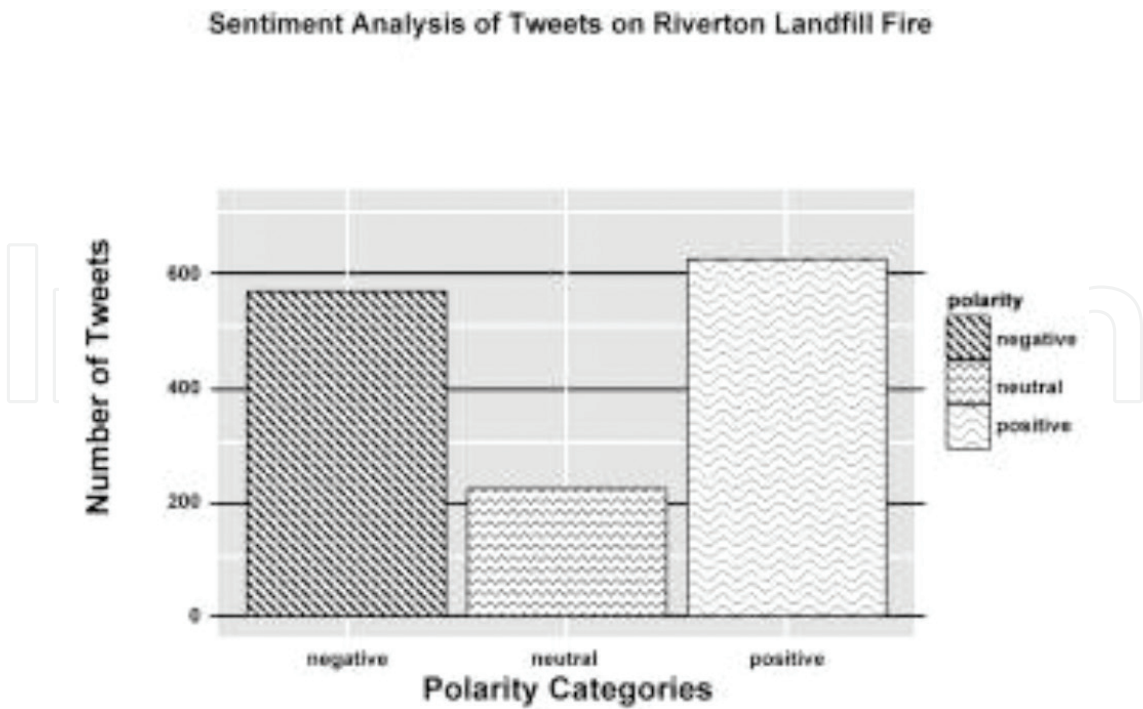
Sentiment Analysis of Tweets on Riverton Landfill Fire

Figure 6. Sentiment polarity of tweets obtained on the Riverton Landfill fire in Jamaica.

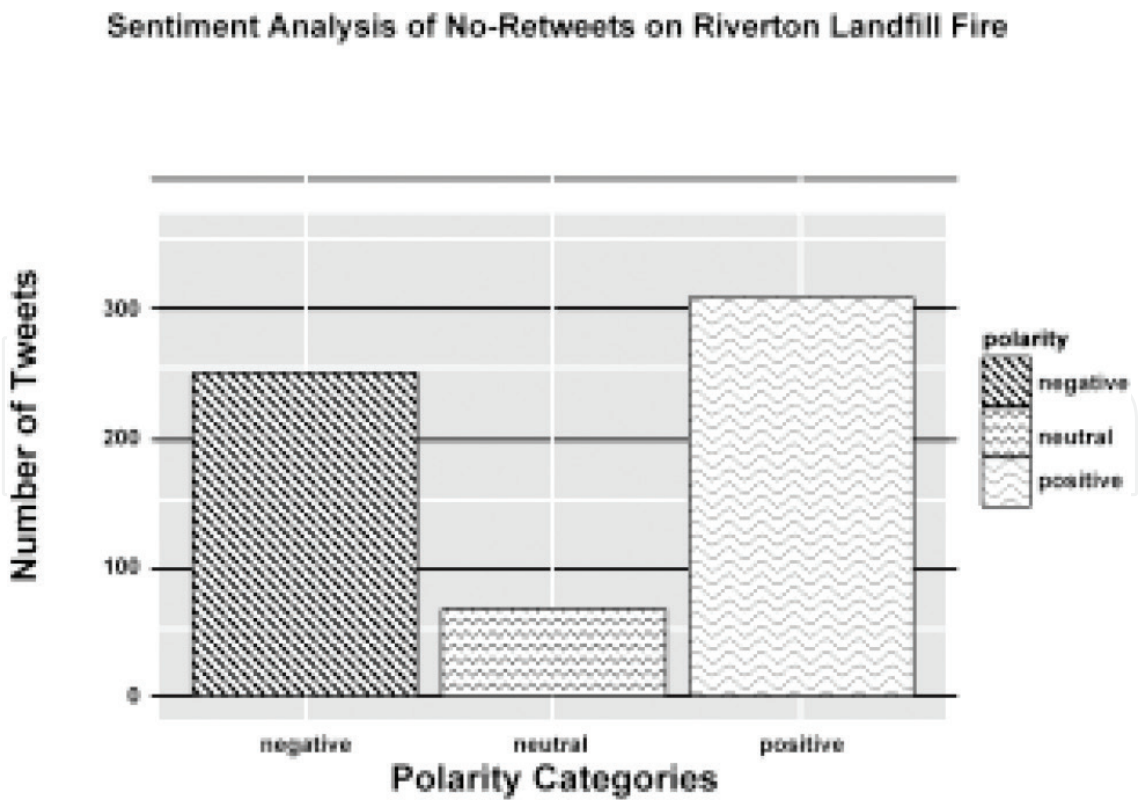## Sentiment Analysis of No-Retweets on Riverton Landfill Fire

Figure 7. Sentiment polarity of no-retweets obtained on the Riverton Landfill fire in Jamaica.

3.1.3.2. Retweets

Of the retweets, three hundred and eleven (311) tweets were negative and three hundred and twenty eight (328) were positive.

3.1.4. Barack Obama's visit to Jamaica

Figure 8 shows that the majority (2583) tweets of the three thousand five hundred and one (3500) tweets collected on Barack Obama's visit to Jamaica were positive. This demonstrates that the Jamaican citizens on Twitter social supported the visit of the President to Jamaica. However, six hundred and fifty eight (658) Jamaicans on Twitter expressed negative sentiments toward Barack Obama's visit to Jamaica. There was a movement suggesting the success of visit of his visit was dependent on whether he offered or announced a Presidential Pardon to the country's first national hero Marcus Garvey, civil rights activist in Jamaica and the USA, who allegedly was falsely convicted of mail fraud in the USA. Failure to grant a pardon to the civil rights activists spurred some of the negative tweets. Hence, *anger* appears on the word cloud in **Figure 9**, generated by RStudio.
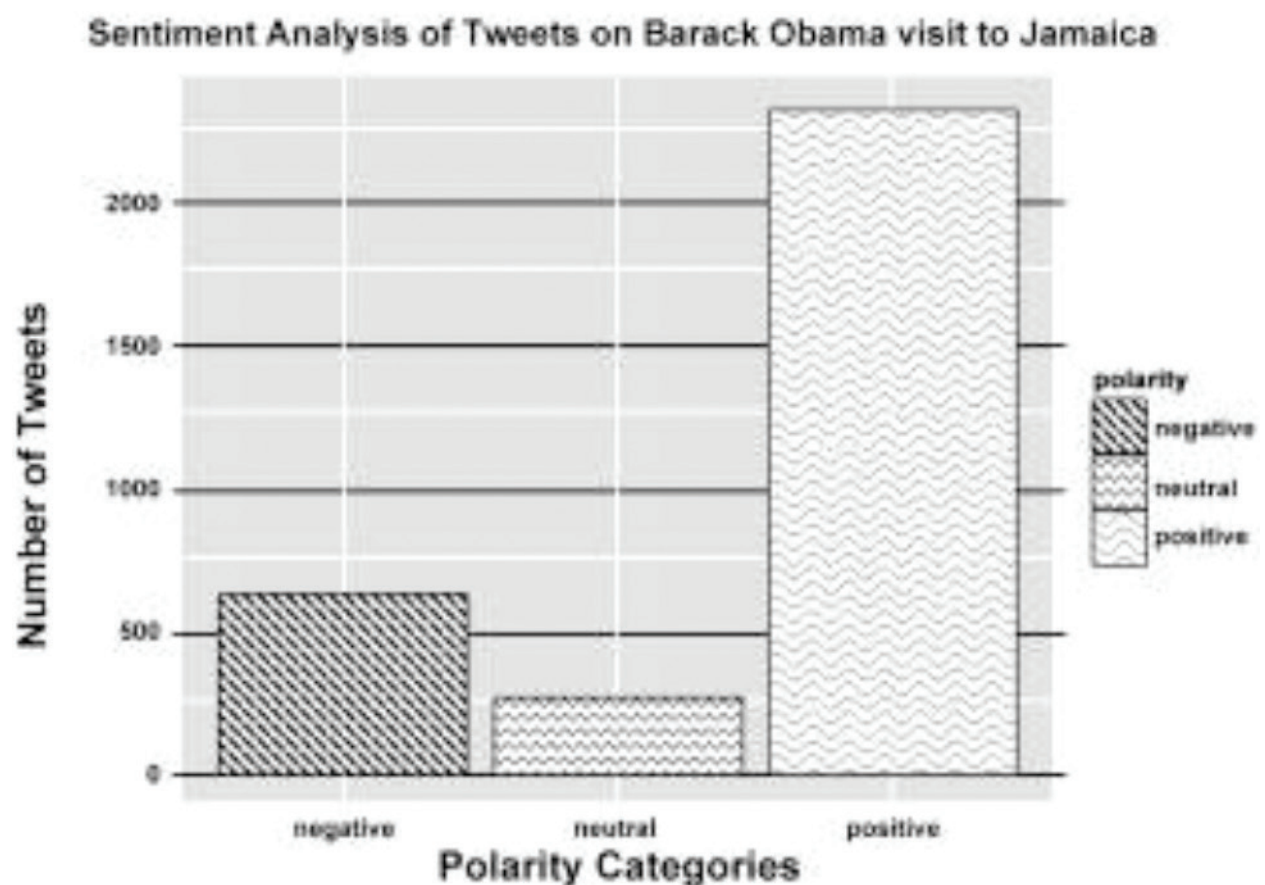
Figure 8. Sentiment polarity of tweets obtained on Barack Obama's visit to Jamaica.
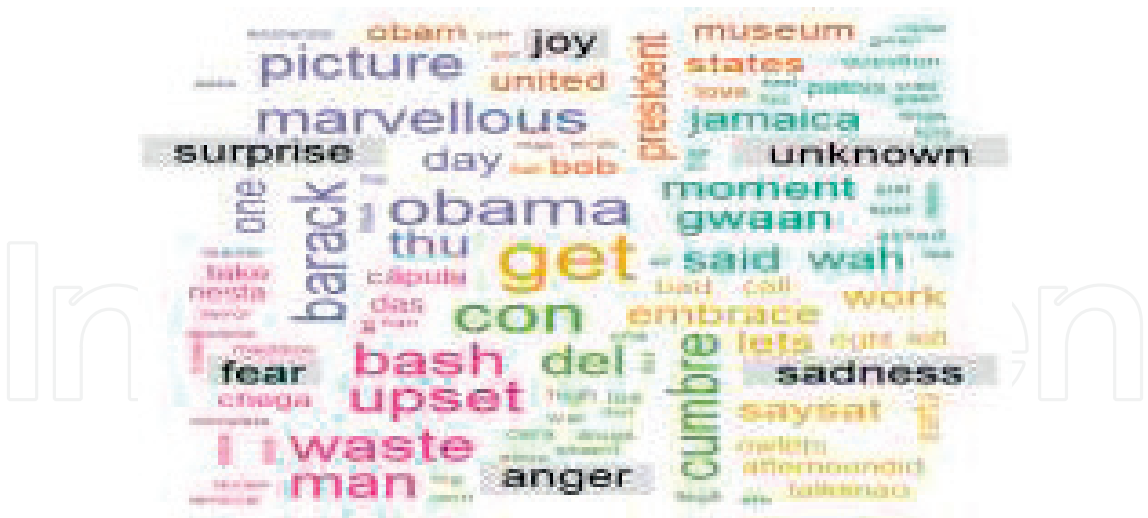
Figure 9. Word cloud showing frequently tweeted words associated with Barack Obama's visit to Jamaica.

3.1.4.1. No-retweets

Figure 10 depicts the results obtained from analysis of tweets that were not retweeted, as in these tweets were posted by the original author. It shows that nine hundred and seventy one (971) tweets were positive and three hundred and twenty two (322) were negative.

3.1.4.2. Retweets

From the analysis of tweets that were retweeted by users who shared similar sentiments of tweets posted by other Twitter users there were three hundred and thirty seven (337) tweets were negative and one thousand six hundred and twelve (1612) were positive.
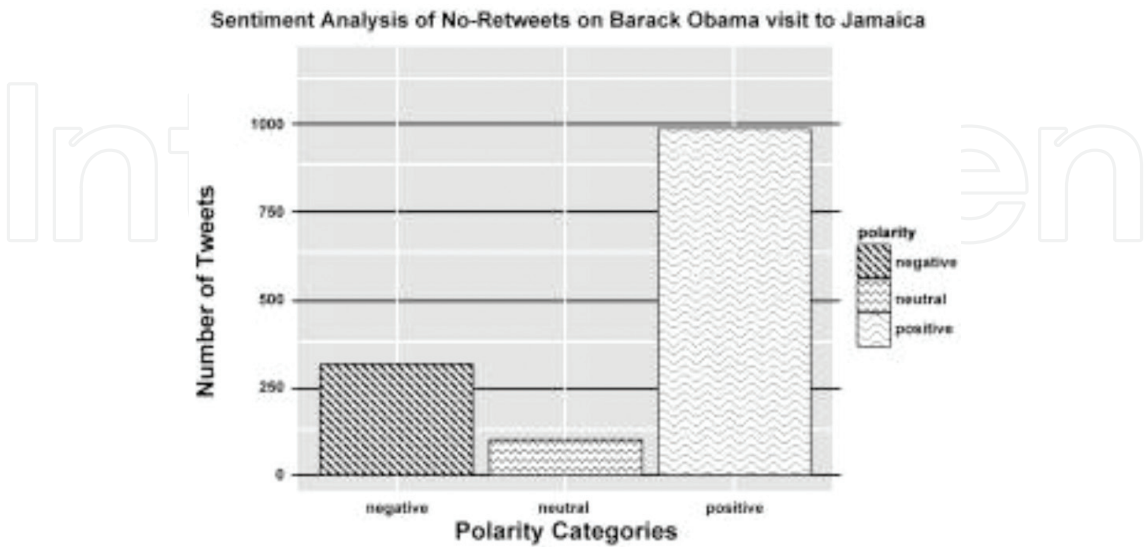

Figure 10. Sentiment polarity of no-retweets obtained on Barack Obama's visit to Jamaica.

3.2. Emotions of topical issues

This section will present information on emotions expressed for each topic, with no-retweets. In everyday speech, emotion is viewed as one's state of mind and instinctive response and are intertwined with mood, temperament, personality and disposition. Emotions are elicited by significant events that are significant when they touch upon one or more of the concerns of the subject. Emotions thus result from the interaction of an event's actual or anticipated consequences and the subject's concerns [6]. In this research several emotions were highlighted: anger, fear, joy, sadness, surprise and disgust. However, due to RStudio's incapability to classify some of the tweets into emotions many tweets were classifieds "unknown".

3.2.1. No-retweets

Tweets posted by authors were of mixed emotions, varying from anger to joy. As depicted in the **Figures 11–14**, RStudio encountered difficulty in classifying the emotions associated with majority of the tweets. As a result of this, majority of the tweets for the decriminalization of marijuana in Jamaica were classified as "unknown" in **Figure 11**. Many factors including the use of the Jamaican creole and the use of sarcasm may have contributed to R's difficulty in determining the emotions of the tweets. This was noticed for the emotions depicted on the other topical issues selected (**Figure 11**).
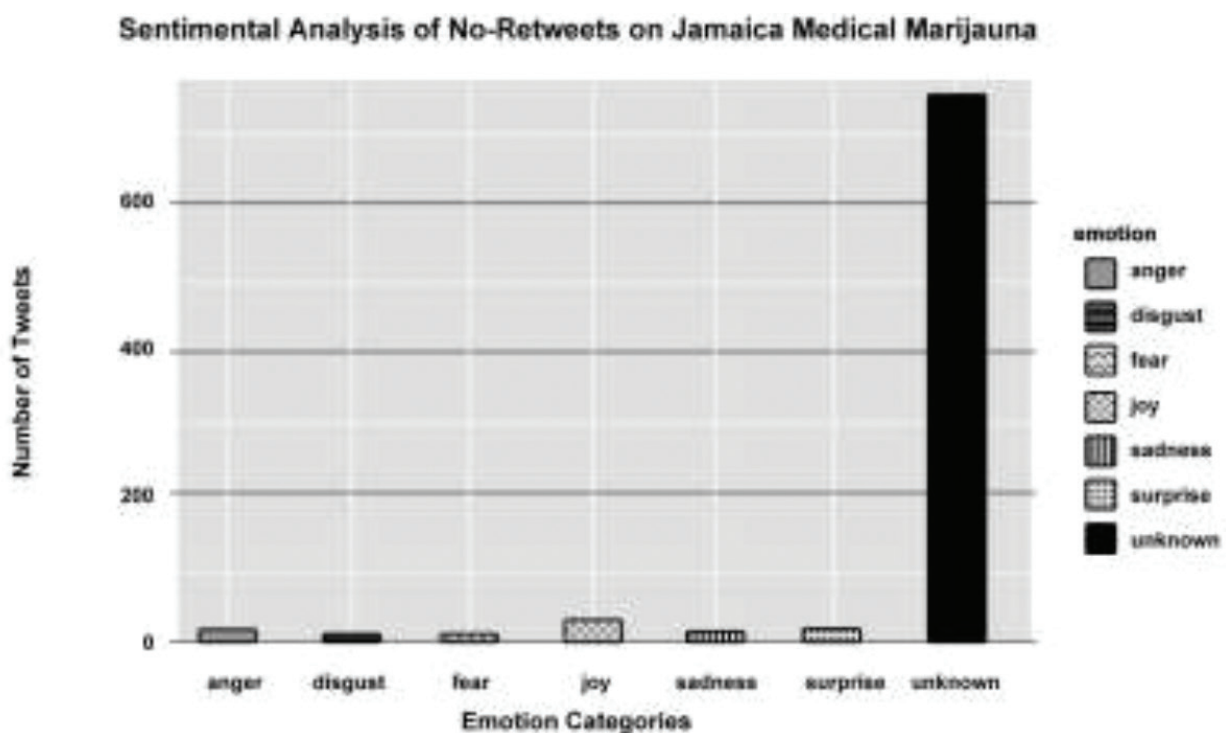
Figure 11. Sentiment emotion of no-retweets obtained on the decriminalization of marijuana in Jamaica.

3.2.2. Retweets

The tweets that were retweeted by Twitter users for the decriminalization of marijuana in Jamaica were of joy and sadness.

3.3. Kaci Fennell's placing in Miss Universe

3.3.1. The decriminalization of marijuana in Jamaica

In **Figure 12**, apart from the tweets that were classified as unknown, it can be seen that tweets expressing anger, joy and surprise recorded the highest numbers. Tweets posted by authors were of mixed emotions, with joy and anger representing the more frequent emotions expressed.

3.3.2. Retweets

The tweets that were retweeted by Twitter users for Kaci Fennell's placing in the Miss Universe event were mostly of joy and anger.

3.4. Riverton Landfill fire

As depicted in **Figure 13**, the emotions discovered for the Riverton Landfill fire varied, more so than the other topical issues that was selected. The tweets analyzed resulted in emotions of anger, disgust, joy, sadness and surprise.
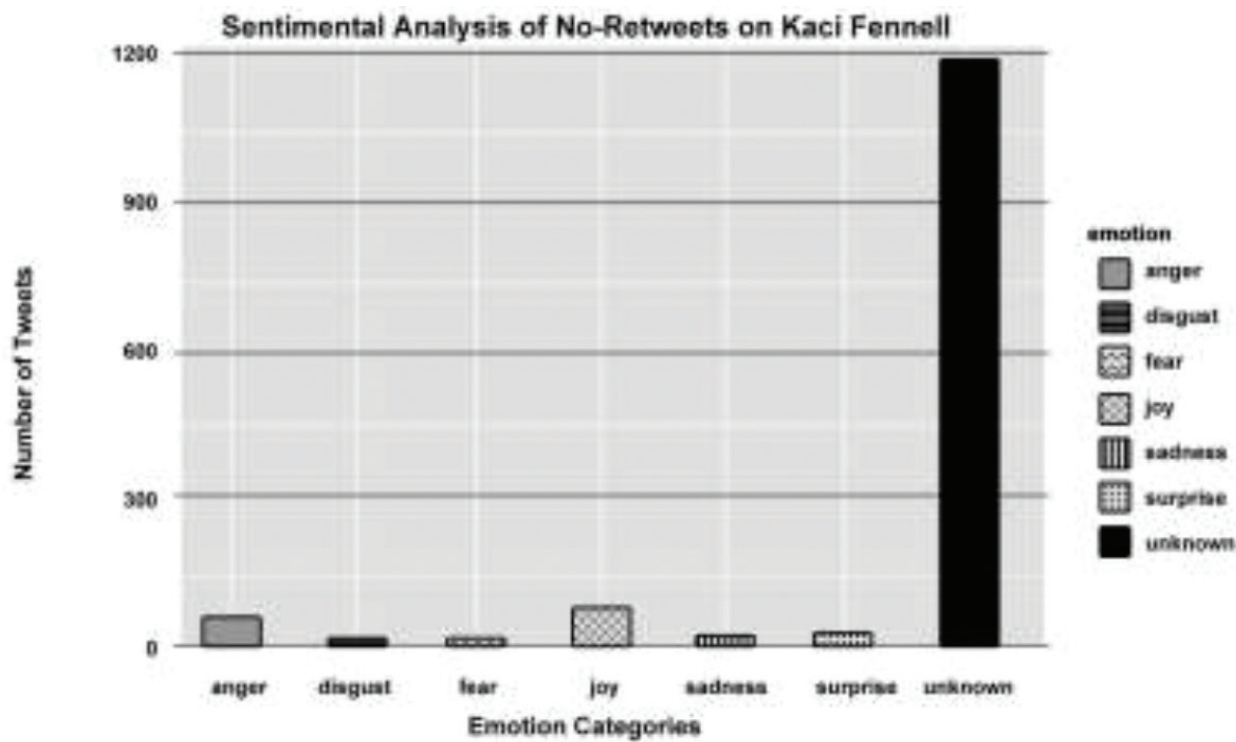

Figure 12. Sentiment emotion of no-retweets obtained on the Kaci Fennell's placing in Miss Universe.
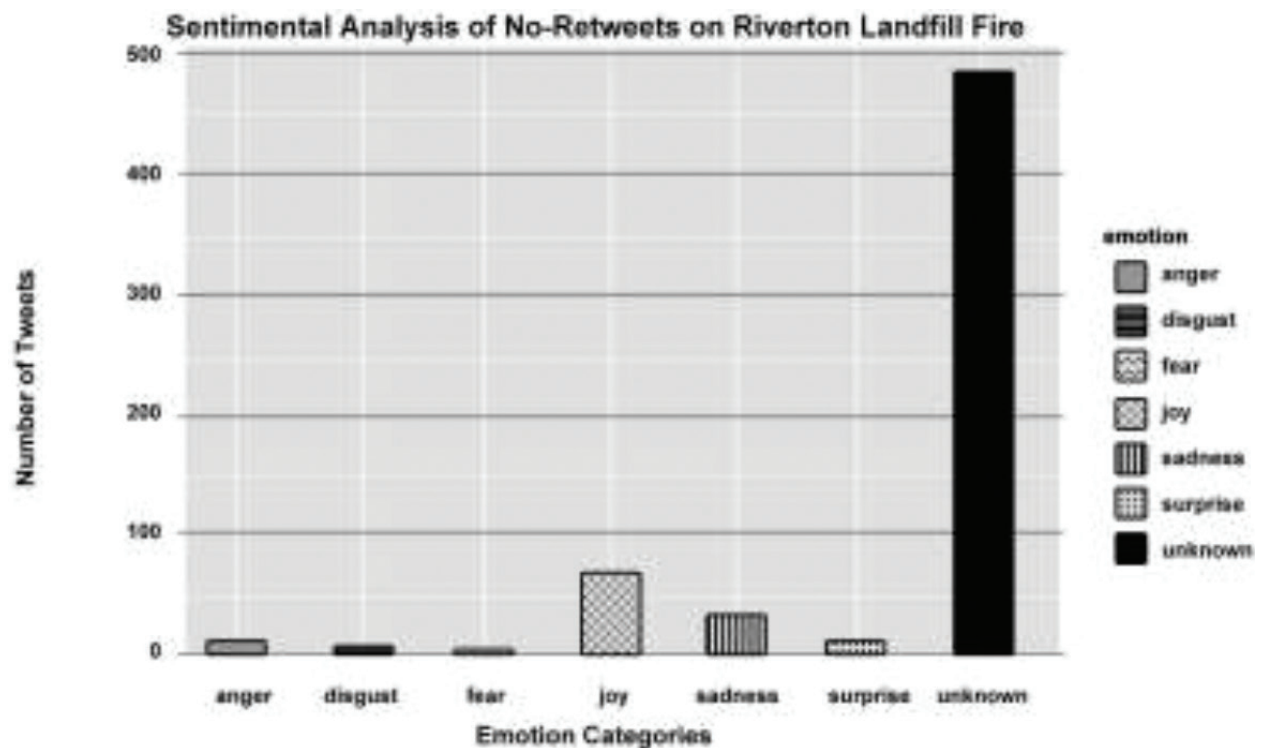
Figure 13. Sentiment emotion of no-tweets obtained on the Riverton Landfill fire.

3.4.1. No-retweets

Tweets posted by authors were of mixed emotions, with joy and sadness representing the more frequent emotions expressed.

3.4.2. Retweets

The tweets that were retweeted by Twitter users on the Riverton Landfill fire were mostly of joy, sadness, disgust and anger.

3.5. Barack Obama's visit to Jamaica

As shown in **Figure 14**, there was difficulty in classifying the emotions associated with majority of the tweets. As a result of this, majority of the tweets for the Barack Obama's visit to Jamaica were classified as "unknown". Many factors including the use of the Jamaican creole and the use of sarcasm may have contributed to R's difficulty in determining the emotions of the tweets. This was noticed for the emotions depicted on the other topical issues selected. Other emotions expressed were of joy and surprise.

3.5.1. No-retweets

Tweets posted by authors were of mixed emotions, varying from joy to surprise to fear.
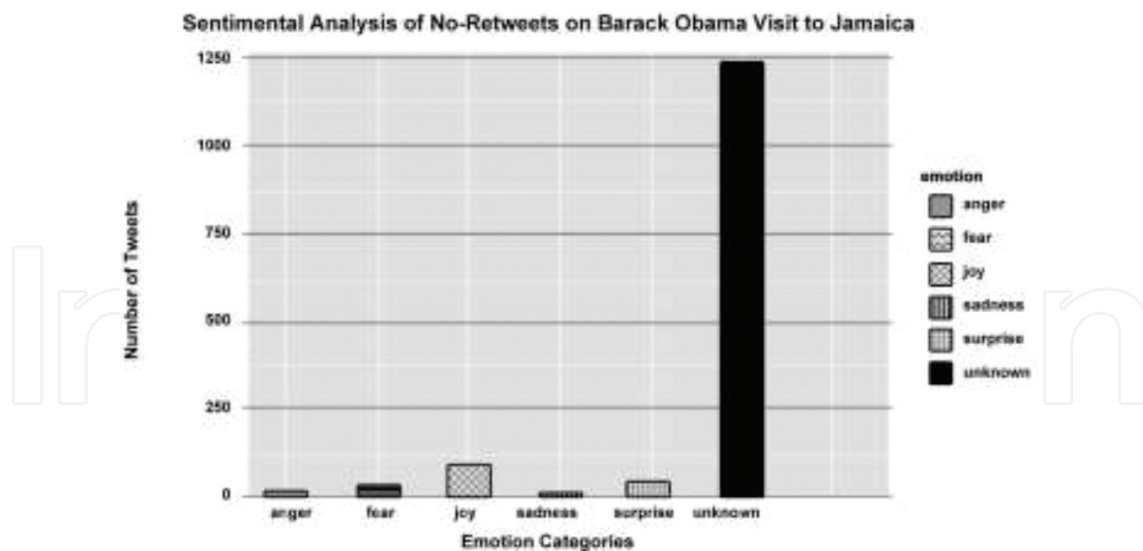
Figure 14. Sentiment emotion of no-retweets obtained on Barack Obama's visit to Jamaica.

3.5.2. Retweets

The tweets that were retweeted by Twitter users for Barack Obama's visit to Jamaica were of joy and surprise.

Sample tweets: This section shows samples of tweets and their sentiment classification for each of the four topical issues (**Tables 1–4**).

Decriminalization of marijuana in Jamaica	
Tweet	**Polarity**
"Jamaica passes law that decriminalizes small amounts of pot. legislation also creates licensing agency to regulate medical"	Positive
"Jamaica decriminalizes marijuana, reminds rest of world it isn't legalized"	Neutral
"in other words, don't get too crazy. weed wasn't legal before, and now it's just less illegal."	Negative

Table 1. Examples of Twitter posts with expressed opinions on the decriminalization of marijuana in Jamaica.

Kaci Fennell's placing in Miss Universe	
Tweet	**Polarity**
"we continue to be proud of kaci fennel"	Positive
"miss jamaica universe kaci fennell will play mass with tribe for carnival 2015 come monday and Tuesday"	Neutral
"miss jamaica kaci fennell 'robbed' of miss universe crown"	Negative

Table 2. Examples of Twitter posts with expressed opinions on Kaci Fennell's placing in Miss Universe.

Riverton Landfill fire	
Tweet	**Polarity**
"said fire would be out by weekend it's not yet Friday"	Positive
"adding to the confusion gleaner when will riverton dump fire be extinguished odpem heads give conflicting deadlines"	Neutral
"a number of schools closed early again today because of rivertondump smoke incl hydel schools amp st patricks primary"	Negative

Table 3. Examples of sample Twitter posts with expressed opinions on the Riverton Landfill fire.

Barack Obama's visit to Jamaica	
Tweet	**Polarity**
"that moment after barack obama said wah gwaan jamaica"	Positive
"barack obama is in jamaica hes just said this"	Neutral
"chronixx upset at barack obamas jamaica visit calls him a waste man bash government"	Negative

Table 4. Examples of Twitter posts with expressed opinions on President Barack Obama's visit to Jamaica.

4. Conclusion

This chapter presented information on how sentiment analysis can be used to extract subjective information from a social media website such as Twitter. It provides researchers with an opportunity to collect deep, rich, readily available qualitative information from a large group of participants in an unobstructed real world environment. Despite several potential uses of sentiment analysis, the literature highlights some general challenges that can be faced when using it with Twitter, such as the noisy nature of Twitter's one hundred and forty (140) character long expressions.

4.1. Significance of the research

Many organizations have taken the initiative to use the tools available through sentiment analysis because of the benefits. The sentiment analysis approach presented in this chapter can be very useful for entities and organizations interested in gathering and understanding the opinions of stakeholders who are Twitter users. This is further facilitated by the availability of Twitter data and posts through Twitter's privacy policy. This study is significant as it presents the results of topical issues being discussed in the country to the public.

This method can be used by companies for marketing research, to aid campaigns and allow stakeholders to understand customer perceptions and thus improve service delivery. Sentiments derived from citizens tweets has even used in forecasting stocks. In Ref. [7], the sentiments marijuana tweets were used to predict stock prices of pharmaceutical companies.

4.2. Recommendations for further research

Overall, the tools available to conduct sentiment analysis on social media sites including Twitter are readily available, but had limitations when used in a Jamaican context. The results received were incorrect at times, which could be as a result of RStudio's inability to understand the Jamaican dialect (patois) that was used in some of the tweets and even the use of sarcasm in some of the tweets presented.

To improve the classification of tweets, in terms of classifying tweets into polarity and emotion, a dictionary of Jamaican words and expressions could be created and then included in the RStudio application. Through the use of this dictionary, in addition to the dictionary already included in RStudio, classification will be improved and misclassification will be deterred when sentiment analysis is used in the Jamaican context.

4.3. Summary

Results indicate that the opinions of Jamaicans on Twitter varied and that many Jamaicans shared the sentiments of others, evidenced by the number of retweets discovered. Among three classification algorithms used, J48 received highest accuracy for the four topics tested and maintained the lowest error rate. The accuracy was just over 70% and the mean absolute error (MAE) was less than 0.3 for the decriminalization of marijuana, Kaci Fennell's placing in the Miss Universe competition and President Obama's visit. For Riverton Landfill fire, the MAE was higher at 0.38 with a comparatively lower accuracy of 55% and precision of 61%. For the decriminalization of Marijuana 72% of the tweets analyzed were positive while for Kaci Fennell's placing in the Miss Universe event 71% of tweets analyzed were of negative sentiments. There was a marginal difference between positive and negative views obtained on the Riverton Landfill fire. Finally, 73% of the tweets collected on President Obama's visit to Jamaica showed positive sentiments, which can be interpreted that many Jamaicans were appreciative of his visit to the island.

Acknowledgements

Author details

Sherrene Bogle

Address all correspondence to: sbogle@utech.edu.jm; sherrene.bogle@gmail.com

University of Technology, Jamaica, Kingston, Jamaica

References

[1] Kennedy H. Perspectives on sentiment analysis. Journal of Broadcasting & Electronic Media. 2012 Oct 1;**56**(4):435-450

[2] Jamaica Information Service. President Barack Obama Arrives in Jamaica [Internet]. 2015. Available from: http://jis.gov.jm/president-barack-obama-arrives-jamaica/

[3] Bogle S, Bogle V, Anderson T. Sentiment analysis of consumers' perceptions on social media about the main mobile providers in Jamaica. World Academy of Science, Engineering and Technology, International Journal of Biomedical and Biological Engineering. 2016;**2**(1):355

[4] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics. Vancouver, Canada. 2005 Oct 6. pp. 347-354

[5] Frijda NH, Manstead AS, Bem S, editors. Emotions and Beliefs: How Feelings Influence Thoughts. Cambridge University Press; 2000 Oct 12

[6] March Fire at Riverton Dump [Internet]. 2015. Available from: http://m.jamaicaobserver. com/news/Report--March-fire-at-Riverton-dump-most-detrimental-in-history_18885503 [Accessed: 11-09-2017]

[7] Bogle SA, Potter WD. SentAMaL-a sentiment analysis machine learning stock predictive model. In: Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 2015. p. 610. ISBN: 1601324057