A Novel Position-based Sentiment Classification Algorithm for Facebook Comments

Khunishkah Surroop
Computer Science & Engineering Department
Faculty of Engineering
University of Mauritius
Reduit, Mauritius

Khushboo Canoo
Computer Science and Engineering Department
Faculty of Engineering
University of Mauritius
Reduit, Mauritius

Sameerchand Pudaruth*
Ocean Engineering & ICT Department
Faculty of Ocean Studies
University of Mauritius
Reduit, Mauritius

***Abstract*—With the popularization of social networks, people are now more at ease to share their thoughts, ideas, opinions and views about all kinds of topics on public platforms. Millions of users are connected each day on social networks and they often contribute to online crimes by their comments or posts through cyber bullying, identity theft, online blackmailing, etc. Mauritius has also registered a surge in the number of cybercrime cases during the past decade. In this study, a trilingual dataset of 1031 comments was extracted from public pages on Facebook. This dataset was manually categorized into four different sentiment classes: positive, negative, very negative and neutral, using a novel sentiment classification algorithm. Out of these 1031 comments, it was found that 97.8% of the very negative sentiments, 70.7% of the negative sentiments and 77.0% of the positive sentiments were correctly extracted. Despite the added complexity of our dataset, the accuracy of our system is slightly better than similar works in the field. The accuracy of the lexicon-based approach was also much higher than when we used machine learning techniques. The outcome of this research work can be used by the Mauritius Police Force to track down potential cases of cybercrime on social networks. Decisive actions can then be implemented in time.**

***Keywords*—sentiment analysis; Facebook; cybercrime; emoticons**

I. INTRODUCTION

The 21st century evokes the magical era of technological advancements amongst which are the evolution of social media sites. People share their thoughts, ideas, opinions, views, knowledge and experiences on platforms such as blogs, social networks, news portals, travel sites and wikis.

"Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes." [1]. Thus, sentiment analysis is a technique that focuses on the detection of favourable and unfavourable opinions about specific subjects. Sentiment analysis usually involves the extraction of sentiments hidden in users' public texts which they publish on online platforms. In recent years, we have seen people sharing their opinions in diverse fields such as marketing, politics, religion, books, movies, sports, health, etc. This increase in online activities have also led to a consequential rise in the number of scams, cyber bullying, cyberagression, blackmails, identity theft, promotion of terrorism and cyber harassment cases. The words and expressions used by users can reveal their intention (sentiment) and therefore necessary measures can be taken to reduce the impact of negative comments. Sentiment Analysis can help to recognize people's emotions and display the polarity of the comments and help in the making of safer online platforms.

The viral power of online media has proven that threats can be spread within seconds. For example, if anyone is publicly criticizing or judging any other religion, these comments and responses can spread within minutes and after only a very short while can lead to cyber-aggression between Face bookers. As a consequence, these comments could rapidly and easily create social instability in the country. Facebook has become the ideal platform to commit all types of cybercrimes.

Bullying has also taken new dimensions as it has become so easy to create a fake account on Facebook, target a victim, update your status and wait for the awful comments that even people unknown to the victim would be posting about him. This may induce serious damage to the latter's mental health. For example, a boy who was cyberbullied died as the insults which he received caused a strong emotional stress which he could not digest [2]. On the 5th of September 2015, Mauritians had witnessed a vague of tension that befell over the country due to a racial riot [3]. This particular case was further dramatized by some surfers against racial and religious harmony who opinioned about the matter in a negative way that intimidated people from different ethnicities and this could cause a serious civil conflict in the country.

There are many pending cybercrime cases that the investigators of the Cybercrime Unit (CCU) of Mauritius are scrutinizing [4]. A list of some of the recently reported cases of cybercrime in Mauritius has been provided in Appendix A. Despite the fact that many cases of cybercrime are identified and reported to the Cybercrime Unit, it is a hectic job for investigators to read all the comments, sometimes in thousands and analyze one by one to find potential offenders.

Thus, in this work, we have developed a sentiment analysis tool for the detection of potential threats prevailing on Mauritian Facebook pages. This tool enabled us to process all posts and comments that are extracted from specific pages on Facebook. An analysis of those texts has been done by filtering the useful keywords that could determine the sentiment of the

text and neglect those that are not important. And finally, a classification is done whereby the sentiment of the text is classified as either positive, negative, very negative or neutral.

The language of the texts has also been taken into consideration. They were classified as Creole, English or French. Face bookers also express their emotions by means of animated facial expressions which are known as emoticons. The value of these emoticons has also been assessed. Furthermore, our tool can save the investigators from long hours of hectic work of reading and analyzing thousands of comments which are posted daily on public pages in Facebook. Instead, this tool will highlight only those comments which require immediate attention and where potentially immediate actions may be taken.

The rest of the paper is discussed as follows: Section II focuses on the previous work carried out by researches worldwide. Our proposed approach is described in Section III. Section IV evaluates the tool based on the manual classification. Finally, Section V concludes the study and mentions the future works.

II. Related Works

Sentiment analysis has been an active field of research since the last decade. However, with the unprecedented growth in the amount of unstructured texts that is being generated online, the field of sentiment analysis is gaining more popularity and is increasingly becoming more important for decision making in large businesses, governmental organizations and many others. However, the increasing usage of social networks has also brought with them new types of social problems and many of them can be categorized as cybercrimes.

Dinakar et al. (2011) emphasized the importance of tracking those harmful comments communicated on the web. They made use of 4500 YouTube comments and classified the data into three areas of bullying based on intelligence, race and culture and sexuality. An overall accuracy of 76% was obtained by using the SMO (Sequential Minimal Optimization) technique [5]. An interesting study by Gerber (2014) analyzed tweets tagged with spatio-temporal information in order to predict crime. Gerber found that the proposed technique was able to improve crime detection for 19 out of the 25 crime types he studied [6].

Bolla (2014) used sentiment analysis techniques to demonstrate that crimes can be detected almost in real-time from social media. He analyzed one hundred thousand tweets from different cities in the United States to determine the crime intensity in each region. It turned out that there was a high correlation with real-life events [7]. Lin (2014) analyzed 180 million tweets after the Boston Marathon bombing events in order to understand how twitter users from 25 major cities expressed fear and/or comfort. He concluded by suggesting that his system could be used to predict perceived threats in the event of natural or man-made disasters [8].

Chen et al. (2012) used comments made on YouTube videos to detect offensive language. They used a combination of lexical and syntactic features to improve the performance of their classifier. They also analysed the writing styles of a sample of users in order to predict their likeliness to send out offensive content. More recently, Krishna (2014) used comments associated with YouTube videos to demonstrate that it is possible to correlate users' sentiments with real-world events especially for popular events or personalities [9][10].

Qin Li (2015) analyzed the accuracy of the sentiment classifiers provided by four different companies involved in online brand monitoring. He found that most of the commercial classifiers cannot deal with negation, emotions and noisy language [11]. Jurek et al. (2015) smart sentiment classification algorithm considers both the presence of negation words and intensity modifiers. They proposed a new approach to determine the overall polarity of a comment when both positive and negative words are present in that comment. However, their approach did not bring significant improvements and can be used only for short sentences. In a previous work, they had used this tool to estimate the level of public disorder in public events by analysing sentiments in tweets [12][13].

Hosseinmardi et al. (2015) analyzed text comments and images from Instagram to detect incidents of cyberagression and cyber bullying using machine learning techniques such as support vector machines and logistic regression. They obtained a recall of 79% and a precision of 71% from text comments. For non-text features, the recall was slightly lower at 76% and the precision at 62% [14].

Xu et al. (2012) have detected traces of bullying on the Twitter micro-blogging platform. Their tasks consisted of obtaining traces via the twitter streaming API to find an instance of the word "bully" in tweets and build an eight classes text classifier based on pre-defined emotion classes [15]. Henri et al. (2012) showed that it is possible to predict real-world threats by extracting abnormalities in tweets [16]. Zhang et al. (2011) predicted text sentiments by using machine learning method to build a web-based system called SES. They conducted their experiments on Facebook comments and tweets using four different machine learning models [17].

While previous works have concentrated on finding cyber bullying instances from only one social network, Dadvar et al. (2012) suggest that using information collected from multiple social networks can improve the tracking and prediction of cyber bullying. They have also incorporated gender as a feature which they believe to have increased the defective accuracy. Their main dataset was collected from the MySpace social network [18]. Via a carefully designed Facebook experiment, Anderson et al. (2014) demonstrated that if a third-party provides support to a victim of cyber bullying through a dissenting comment, this encourages other people to provide support and empathy for the victim [19].

While many studies mentioned about how to detect instances of cybercrime, most of them did not explain how this information can be used. Cohen et al. (2014) proposed a safer internet by identifying cyber bullying from Facebook comments and trying to mitigate it by informing the relevant institutions in order to provide assistance to those being bullied. They understood that while technology has made bullying easier and faster, they also proposed that the same technology be used to detect and report such offences [20].

From the analysis of about 7300 tweets related to cyber bullying, Alim (2015) found that a significant minority of users are using tweets for reporting cyber bullying cases and for providing advice to victims, parents or school administrators [21]. Duwairi et al. (2014) has developed a framework to analyze tweets in Arabic dialects and Arabizi. Their proposed framework can handle repetitions, emoticons and negation. The highest accuracy achieved was 76.8% when using the Naïve Bayes classifier. Lexicon expansion through the use of synonym has a drawback of the wording losing its primary meaning after a few recapitulations [22][23].

Troussas et al. (2014) classified Facebook status messages into positive and negative sentiments using three different classifiers. They found that the Naïve Bayes classifier had the highest precision (77%) while the Rocchio classifier had the highest recall (73%) and both classifiers outperformed the Perceptron classifier [24].

In order to detect changes in emotions in Facebook messages, Ortigosa et al. (2013) used a combination of machine learning techniques and hand-crafted lexicons to achieve an overall accuracy of 85% [25]. Agarwal et al. (2011) made use of parts-of-speech (POS) features on tweets which allowed them to use a smaller set of features to achieve similar level of performances as reported in similar works [26]. However, Kouloumpis et al. (2011) concluded that POS features may not be useful at all for classification of tweets [27].

There have been multiple previous works dealing with the number of likes, types of emoticons used or hash tags to predict the emotions or opinions of people based on datasets extracted from Twitter and Facebook. However, these experiments were limited to sentiment classification using the English dictionary only, rather than dealing with sentiments in a multilingual setting dealing with cybercrime. Moreover, there is no research that has been carried out in Mauritius to deal with issues of cybercrimes in the Creole language.

III. Methodology

Our approach consists of three steps namely data extraction, data processing and data classification. In addition to this, we have catered for colloquial words written differently by different users.

A. Data Extraction

The dataset was extracted using the Facepager tool [29]. Facepager is a tool for fetching data that is publicly available from Facebook and Twitter. All the extracted data are saved in an SQLite database and may also be exported to a csv file. Our dataset consists of comments extracted from different posts including videos, pictures and links. The trilingual dataset consists of 1031 comments each having English, French and Creole words together.

B. Data Processing

The next step is to process and clean the data. The raw data is practically unreadable due to the presence of extra information such as punctuations and other symbols. Tokenization is first applied to break the sentences into distinct words. In the process, emoticons have also been captured. Our work comprised of finding a link between what people write and what their smileys tend to depict.

1) Emoticons Lexicon

Our tool captured emoticons and displayed them as texts. We assessed the same post in two different desktop computing platforms (Microsoft Windows and Ubuntu) and two different mobile platforms (iOS and Android). It was found that comments from iOS and Android mobile phones generated the same kind of symbols and symbols from the desktop platforms were also similar but were different from each other. However, we noticed that there is a lack of correlation between the sentiments expressed in the text of a comment and the emoticons that are associated with it. Thus, these emoticons are not considered further in this paper.

TABLE I. Emoticons from Different Computing Platforms

Emoticon	Sentiment	Windows/Ubuntu	iOS/Android
	Smile	**:-)**	**ðŸ˜Š**
	Sad	**:-(**	**ðŸ˜’**
	Cry	**:'(**	**ðŸ˜¢**
	Kiss	**:***	**ðŸ˜˜**
	Confused	**:/**	**ðŸ˜**
	Laughing	**:D**	**ðŸ˜„**

2) Dictionaries

We made use of six dictionaries namely the Creole dictionary containing 3063 Creole words, the English dictionary containing 110 206 English words, the French dictionary containing 336 534 French words, the Positive Words dictionary containing 977 words, the Negative Words dictionary containing 1344 words and the Very Negative Words dictionary containing 200 words. These 200 words are highly offensive and cannot be ignored even if they are preceded by a negation word. Examples of such words are: terrorism, kill, drugs, etc. The last three dictionaries contain words in all the three languages.

3) Language Classification Algorithm

The trilingual dataset consisted of comments having English, French and Creole words extracted from Facebook pages in Mauritius such as L'Express, Le Défi Media and IslandCrisis (Appendix A).

The Language Classification algorithm work as follows. If 30% of the words in a post are in the Creole language or if the number of Creole words is greater or equal to the number of English and French words, the language of the post is classified as Creole. This percentage is calculated by taking the number of Creole words in the post and dividing it by the total number of words in the post. If the number of French words is greater than the number of Creole and English words, the language is classified as French. Finally, if the number of English words is greater than the number of French and Creole words, the

language is classified as English otherwise the language is classified as Other. An example of a post is provided below.

"Its dumb zat an IT specialist wud do such a thing zat 2 frm his own cyber cafe!!Someone dignity is at stake. Eski gouvermen pou dedommage li si li inocen!!ki pu ariv so réputation?"

Creole words: '*Eski*', '*gouvermen*', '*pou*', '*li*', '*si*', '*li*', '*inocen*', '*ki*', '*pu*', '*ariv*', '*so*'.

English words: '*Its*', '*dumb*', '*zat*', '*an*', '*specialist*', '*wud*', '*do*', '*such*', '*thing*', '*zat*', '*frm*', '*his*', '*own*', '*cyber*', '*Someone*', '*dignity*', '*is*', '*at*', '*stake*'.

French words: '*cafe*', '*réputation*', '*dedommage*'.

There are 11 Creole words, 19 English words and 3 French words in this post. The percentage of Creole words in this post is 33.3% and since this is over 305, the post would be classified as being in the Creole language. Words containing only one character are not considered in the calculation.

4) The Mean Algorithm for Sentiment Classification

The Mean Algorithm starts by determining the position of each positive and negative word in a comment. The Mean Positive Score is calculated by taking the sum of all the position of each positive word and dividing it by the total number of positive words. Similarly, for the Mean Negative Score, the sum of all the negative words indexes is calculated and it is divided by the total number of negative words.

If the Mean Positive Score exceeds the Mean Negative Score, the sentiment will be positive. Furthermore, if the Mean Negative Score exceeds or is equal to the Mean Positive Score, the sentiment will be negative otherwise the sentiment is assumed to be neutral. An example is illustrated below:

"Li ene zoli garson. Li ena talen. mai li movai.", which means, "He is a handsome boy. He has talent but he is bad.", in English.

TABLE II. INDEX OF WORDS IN SENTENCE: MEAN ALGORITHM

Index	Creole Word	English Word	Sentiment
1	Li	He	
2	ene	is	
3	zoli	handsome	positive
4	garson	boy	
5	Li	He	
6	ena	has	
7	talen	talent	positive
8	mai	but	
9	li	he	
10	movai	bad	negative

Applying the mean algorithm, the Mean Positive Score is calculated as (=(3+7)/2) resulting in a score of 5 and the Mean Negative Score is calculated as (=(10)/1) resulting in a score of 10. The sentiment of this sentence is classified as Negative as the Mean Negative Score exceeds the Mean Positive Score.

5) The Adaptive Algorithm for Sentiment Classification

This algorithm works only in the presence of modifiers such as '*not*', '*pa*', etc., in a sentence. These modifiers have the effect of reversing the sentiment of the word immediately following them. This algorithm works in the same way as the Mean Algorithm except that it reverses the polarity of the word if it is preceded by a negation word. For example; "*He is not bad.*" is a positive sentence since '*not*' reversed '*bad*' into a positive word. An example in Creole is illustrated as shown below:

"Li zourer bouku mai selma li pa mover.", which means *"He swears a lot but he is not bad."*, in English.

TABLE III. INDEX OF WORDS IN SENTENCE: ADAPTIVE ALGORITHM

Index	Creole Word	English Word	Sentiment
1	Li	He	
2	zourer	swears	negative
3	bouku	a lot	
4	mai	but	
5	selma	but	
6	li	he	
7	pa	not	
8	mover	bad	negative

This sentence contains the '*pa'* modifier which means '*not'* in English and is followed by a negative word '*mov*er'. Therefore, the modifier reverses the sentiment into a positive one. Using a naive sentiment classifier, this comment would have been classified as negative because the number of negative words exceeds the number of positive words.

Using the mean algorithm, the result will be negative as the Mean Negative Score which is 5 (=(2+8)/2) exceeds the Mean Positive Score which is 0. Using the adaptive algorithm, the Mean Positive Score will be 8 and the Mean Negative Score will be 2 resulting in having an overall positive sentiment, which is a more appropriate classification for this post.

C. Data Classification

After processing the data, sentiments were assigned to 3 different levels: (i) word-level sentiment in which each word in each comment is assigned a polarity, (ii) comment-level sentiment in which each comment from each post is assigned a polarity, and finally, (iii) post-level sentiment whereby each post on choosen pages is assigned a polarity.

The number of positive, negative and very negative words was also calculated during the process. The number of comments in a post was also noted and the number of words categorized under each language was also done. A further level of classification for the *very negative* category has also been implemented. It has been further classified into 7 categories namely Accident, Crime, Cyber bullying, Drug, Racism, Terrorism and Other.

IV. EVALUATION AND DISCUSSION

Three independent annotators were asked to manually classify the data. Because of different responses from the three annotators, it was necessary for us to assign a final sentiment or language to the comments by using the simple majority rule. In a scenario where one annotator classifies a comment as neutral, another one as positive and the third one as negative, the comment is assumed to be negative.

A. Evaluation of Language Classification Results

An attempt was made to classify each comment into an appropriate language. The highest accuracy was obtained for

the English language, possibly because of the big size of the dictionary. The accuracy of the Creole language was also very high. An overall accuracy of 90.9% was obtained for the whole dataset.

TABLE IV. PERCENTAGE ACCURACY FOR EACH LANGUAGE

Language	Manual Categorization	Automatic Classification	% Accuracy
Creole	696	647	93.0
English	220	216	98.2
French	106	74	69.8
All Three	1031	937	90.9

Some comments that were not properly categorised was due to the use of slangs that were not present in our dictionaries. For example, some people had different ways of writing "*good*" so all possibilities could not be captured. A lower accuracy was registered for the French language as many of the French words are very similar to the Creole words and are thus classified into the Creole language.

B. Evaluation of Sentiment Classification Results

In this section, the trilingual dataset has been evaluated on their sentiment classification using the adaptive algorithm.

TABLE V. PERCENTAGE ACCURACY FOR SENTIMENT CLASSIFICATION

Sentiment	Manual Categorization	Automatic Classification	% Accuracy
Negative	362	256	77.0
Positive	300	231	70.7
Very Negative	93	91	97.8
Neutral	276	220	79.7
All Four	1031	798	77.4

It was found that 798 comments were properly classified out of the 1031 comments with an overall accuracy of 77.4 % for the sentiment classification.

An accuracy of 97.8% for the *very negative* category will prove to be highly beneficial for this study as the main aim of our work was to identify potential threats on social networks. A lower accuracy was registered for the *negativ*e sentiment category because of the sarcastic comments that were correctly classified by the human annotators but the algorithm could not recognize them as such.

Sanchez and Kumar (2013) classified Twitter messages using a lexicon of commonly used terms of abuse, which was itself extracted from Twitter. These messages were then classified as either positive or negative. An overall accuracy of 70 % was obtained [28]. Compared to their work, our tool managed to classify comments into four sentiments with a better accuracy of 77.4%. Moreover, we made use of a dataset that consisted of comments written in three different languages. Facebook comments are usually more noisy and lengthy than tweets and people tend to be more informal on Facebook than on Twitter. The adaptive algorithm gave more accurate and precise results than a naive frequency-based sentiment classifier (69.9%) and the mean algorithm (72.4%).

C. Evaluation of Very Negative Sentiments

It was found that 69 comments were probably classified out of the 93 comments. An overall accuracy of 74.2% was obtained.

TABLE VI. ACCURACY OF VERY NEGATIVE SENTIMENTS

Threats	Manual Categorization	Automatic Categorization	% Accuracy
Accident	14	8	57.1
Crime	23	16	69.6
Cyberbullying	5	3	60.0
Drugs	8	7	87.5
Racist	7	6	85.7
Terrorism	27	23	85.2
All Seven	93	69	74.2

The main problem that was encountered was due to the Creole language as it is a language without a universally agreed syntax and grammatical rules. It would not be exaggerated to say that each and very Mauritian has their own unique ways of writing the Creole language. And, therefore, it is very challenging to cater for such a high degree of variability in a language. Furthermore, most Mauritians tend to comment on Facebook in a colloquial manner which sometimes is an amalgam of different languages, all in one sentence only, and this adds another layer of complexity in the classification.

D. Machine Learning Algorithms

To understand the performance and reliability of our proposed algorithm, we also classified the dataset using two well-known machine language algorithms namely the Naïve Bayes and the k-Nearest Neighbour (k-NN) classifiers.

1) Naïve Bayes Classifier

The overall accuracy of the Naïve Bayes classifier for sentiment classification is 45.2%. The highest recall (63.0%) and precision (57.3%) was for the *negative* sentiment. Surprisingly, both the recall and the precision values for the very *negative sentiment* were very low at only 33.3% and 15.4%, respectively. One reason for this is the much lower number of training instances for this category compared to the others. In general, the accuracies for each lower was quite low because the instances from the *neutral* category was uncorrected predicted into the other three categories.

TABLE VII. PERCENTAGE ACCURACY FOR SENTIMENT CLASSIFICATION

Predicted	True Negative	True Positive	True Very Negative	True Neutral	Precision
Negative	228	67	41	62	57.3%
Positive	57	131	13	55	51.2%
Very Negative	48	39	31	83	15.4%
Neutral	29	63	8	76	43.2%
Recall	63.0%	43.7%	33.3%	27.5%	

The overall accuracy of the Naïve Bayes classifier for language classification is 84.6%. The classifier does a much better job at the classification of language than sentiments. The accuracy for the Creole and English are very high (89.1%). The accuracy for the French is probably lower because of the lesser

number of training instances compared to the two others.

TABLE VIII. PERCENTAGE ACCURACY FOR LANGUAGE CLASSIFICATION

Predicted	True Creole	True English	True French	True Other	Precision
Creole	640	31	45	2	89.1%
English	21	172	0	0	89.1%
French	25	9	56	3	60.2%
Other	10	8	5	4	14.8%
Recall	92.0%	78.2%	52.8%	44.4%	

2) K-Nearest Neighbour Classifier

It was found that the overall accuracy for the sentiment classification using the k-Nearest Neighbour (k-NN) was 47.8% which was better than the overall accuracy of the Naïve Bayes classifier. The k-NN classifier also produced much better results for the *very negative* and *neutral* sentiment categries. However, Naïve Bayes outperformed k-NN for the *positive* and *negative* sentiments.

TABLE IX. PERCENTAGE ACCURACY FOR SENTIMENT CLASSIFICATION

Predicted	True Negative	True Positive	True Very Negative	True Neutral	Precision
Negative	165	47	26	66	54.3%
Positive	67	178	16	78	52.5%
Very Negative	26	7	27	9	39.1%
Neutral	104	68	24	123	38.6%
Recall	45.6%	59.3%	29.0%	44.6%	

The overall accuracy for the language categorization was 78.5%. Thus, the k-NN classifier could not match the more accurate results of the Naïve Bayes algorithm. Only the precision for the Creole language and recall for the French language was slightly better.

TABLE X. PERCENTAGE ACCURACY FOR LANGUAGE CLASSIFICATION

Predicted	True Creole	True English	True French	True Other	Precision
Creole	567	34	23	1	90.7%
English	48	166	5	1	75.5%
French	51	8	71	2	53.8%
Other	30	12	7	5	9.3%
Recall	81.5%	75.5%	67.0%	55.6%	

3) Machine Learning v/s The Adaptive Algorithm

In this section, we compare the results obtained from using the adaptive algorithm which we have proposed with two traditional machine learning algorithms. The results are presented in Table XI and Table XII. The precision values have been used for the comparisons.

TABLE XI. PERCENTAGE ACCURACY FOR SENTIMENT CLASSIFICATION

Sentiment	% Accuracy of Novel Algorithm	% Accuracy of Naïve Bayes	% Accuracy of k-NN
Negative	70.7	57.3	54.3
Positive	77.0	51.2	52.5
Very Negative	97.8	15.4	39.1
Neutral	79.7	43.2	38.6
Overall	77.4	45.2	47.8

From Table XI, we observe that the adaptive algorithm outperforms both the Naïve Bayes and the k-NN algorithms in all the four sentiment categories.

TABLE XII. PERCENTAGE ACCURACY FOR LANGUAGE CLASSIFICATION

Sentiment	% Accuracy of Novel Algorithm	% Accuracy of Naïve Bayes	% Accuracy of k-NN
Creole	93.0	89.1	90.7
English	98.2	89.1	75.5
French	69.8	60.2	53.8
Overall	86.8	84.6	78.5

From Table XII, again, we observe that the adaptive algorithm outperforms both the Naïve Bayes and the k-NN algorithms in all the three language categories. Thus, in this paper, we have developed a new polarity assignment technique and we have shown that it performs much better than naïve frequency-based sentiment classifiers and machine learning algorithms. We believe that with more customization of the dictionaries and a larger dataset, it is possible to further improve the overall accuracies of the system.

V. CONCLUSIONS

Due to the rise of cybercrimes, it has become essential for the government to monitor online activities on social networks. Thus, this paper aimed at developing sentiment analyzers to detect potential threats on social networks. After investigating about cybercrime and sentiment analysis related topics, the sentiment analyzer was built using different tools. A novelty aspect of our paper is the sentiment analysis of Creole texts from Facebook, in addition to French and English comments. Another contribution is the analysis of emoticons from different platforms. We have used four different emotion classes compared to most researchers who have used only two. Hierarchical classification of the most serious threats has also been implemented. We also investigated three different algorithms for the classification of comments into an appropriate sentiment class. We also showed that the adaptive algorithm we have proposed produced much better results that machine learning algorithms. This work can be improved by doing the analysis in real-time and communicating the relevant results to the relevant authorities. Using a larger dataset and larger dictionaries will also help to improve the accuracy.

REFERENCES

[1] Liu, B. *Sentiment Analysis: Opinion Mining, Sentiments and Emotions.* Cambridge University Press, 2012.

[2] Lexpress.mu. "Cyberbullying: Akash Callikan porte plainte a la cybercrime unit," *Lexpress.mu*, Mar. 16, 2014.

[3] Hibz, Y. D. "Online Racial Hatred Incitement: Police Elaborated a List of 30 Suspects," *Island Crisis*, Sept. 8, 2015.

[4] Lexpress.mu. "Cybercrime Unit: dans l'univers des enquêteurs du virtuel," *Lexpress.mu*, Oct. 10, 2015.

[5] Dinakar, K., Reichart, R. and Lieberman, H., "Modeling the Detection of Textual Cyberbullying," in *The Social Mobile Web (ICWSM) Workshop*, Barcelona, Spain, 2011, pp. 11-17.

[6] Gerber, M. S., "Predicting Crime Using Twitter and Kernel Density Estimation," *Decision Support Systems*, vol. 61, pp. 115-125, 2014.

[7] Bolla, R. A., "Crime pattern detection using online social media," M.S. thesis. Missouri University of Science and Technology, USA, 2014.

[8] Lin, Y. R., "The Ripples of Fear, Comfort and Community Identity During the Boston Bombings," in *iConference*, Pittsburgh, USA, 2014, pp. 708-720.

[9] Chen, Y., Zhou, Y., Zhu, S., and Xu, H., "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety," in *International Conf. on Social Computing*, Amsterdam, Netherlands, 2012, pp. 71-80.

[10] Krishna, A., "Polarity trend analysis of public sentiment on YouTube," M.S. Thesis, Iowa State University, Iowa, USA, 2014.

[11] Li, Q., "Examining the accuracy of sentiment analysis by brand monitoring companies," in *5th IBA Bachelor Thesis Conf.*, Enschede, The Netherlands, 2015, pp. 1-7.

[12] Jurek, A., Bi, Y., and Mulvenna, M., "Twitter Sentiment Analysis for Security-Related Information Gathering," in *Joint Intelligence & Security Informatics Conf.*, Hague, The Netherlands, 2014, pp. 48-55.

[13] Jurek, A., Mulvenna, M. and Bi, Y., "Improved lexicon-based sentiment analysis for social media analytics," *Security Informatics*, vol. 4(9), pp. 1-13, 2015.

[14] Hosseinmardi, H., Mattson, S. A., Rafiq, R. I., Han, R., Lu, Q. and Mishra, S., "Prediction of Cyberbullying Incidents on the Instagram Social Network," arXiv:1508.06257 [cs.IR].

[15] Xu, J. M., Zhu, X. and Bellmore, A., "Fast Learning for Sentiment Analysis on Bullying," in *International Workshop on Issues of Sentiment Discovery and Opinion Mining*, Beijing, China, 2012.

[16] Henri, B., Olga, R., Daniel, W., Corne, V. and Harry, W., "On the early detection of threats in the real world based on open-source information on the internet," in *International Conf. on Information Technologies and Security (ITSEC)*, Hague, The Netherlands, 2012, pp. 1-12.

[17] Zhang, K., Cheng, Y., Xie, Y., Honbo, D., Agrawal, A., Palsetia, D., Lee, K., Liao, W. nd Choudhary, A., "SES: Sentiment Elicitation System for Social Media Data," in *11th IEEE International Conference on Data Mining Workshops*, Vancouver, USA, 2011, pp. 129-136.

[18] Dadvar, M. and Jong, F.D., "Cyberbullying Detection: A Step Toward a Safer Internet Yard," in *21st International World Wide Web Conf.*, New York, USA, 2012, pp. 121-125.

[19] Anderson, J., Bresnahan, M. and Musatics, C., "Combating Weight-Based Cyberbullying on Facebook with the Dissenter Effect," *Cyberpsychology, Behavior, and Social Networking*, vol. 17(5), pp. 281-286, 2014.

[20] Cohen, R., Lam, D. Y., Agarwal, N., Cormier, M., Jagdev, J., Jin, T., Kukreti, M., Liu, J., Rahim, K., Rawat, R., Sun, W., Wang, D. and Wexler, M., "Using Computer Technology to Address the Problem of Cyberbullying," *Computers & Society*, vol. 44(2), pp. 52-61, 2014.

[21] Alim, S., "Analysis of Tweets Related to Cyberbullying: Exploring Information Diffusion and Advice Available for Cyberbullying Victims," *International Journal of Cyber Behavior, Psychology and Learning*, vol. 5(4), pp. 31-52, 2015.

[22] Duwairi, R. M., Marji, R., Sha'ban, N. and Rushaidat, S., "Sentiment Analysis in Arabic Tweets," in *5th IEEE International Conf. on Information and Communication Systems*, Irbid, Jordan, pp. 1-6, 2014.

[23] Adedoyin-Olowe, M., Gaber, M. M. and Stahl, F., "A Survey of Data Mining Techniques for Social Media Analysis," *Journal of Data Mining & Digital Humanities*, June 2014.

[24] Troussas, C., Virvou, M., Espinosa, K. J., Liaguno, K. and Caro, J., "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," in *4th IEEE International Conf. on Information, Intelligence, Systems and Applications*, Piraeus Greece, 2013, pp. 1-6.

[25] Ortigosa, A., Martin, J. M. and Carro, R. M., "Sentiment analysis in Facebook and its application to e-learning," *Computers in Human Behavior*, vol. 31(1), pp. 527-541, 2014.

[26] Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passoneau, R., "Sentiment Analysis of Twitter Data," in *Proceedings of the workshop on Languages in Social Media*, Stroudsburg, PA, USA, 2011, pp. 30-38.

[27] Kouloumpis, E., Wilson, T. and Moore, J. D., "Twitter Sentiment Analysis: The Good the Bad and the OMG!," in 5th *International Conf. on Weblogs and Social Media*, Barcelona, Spain, 2011, pp. 538-541.

[28] Sanchez, H. and Kumar, S., "Twitter Bullying Detection," University of California, Santa Cruz, California, USA, 2011.

[29] strohne, "Strohne/Facepager," in github, GitHub, 2016. [Online]. Available: https://github.com/strohne/Facepager. Accessed: Oct. 27, 2016.

APPENDIX A: A SAMPLE OF REPORTED CYBERCRIME CASES IN MAURITIUS

Month & Year	Original Title	Title in English	Source
June 2016	Arrêté pour avoir posté des vidéos pornographiques, il est libéré sous caution	Arrested for posting pornographic videos , he was freed on bail.	http://www.lexpress.mu/article/282792/arrete-pour-avoir-poste-videos-pornographiques-il-est-libere-sous-caution
March 2016	Policiers filme: Les commentaires sur facebook font l'objet d'une enquete.	Policemen filmed: Comments on Facebook are subject to an investigation.	http://defimedia.info/policier-filme-les-commentaires-sur-facebook-font-lobjet-dune-enquete-20766/
March 2016	Peut-on tout publier sur Facebook? La Cybercrime Unit demarre une campagne.	Can we all publish on Facebook? The Cybercrime Unit starts a campaign.	http://www.lexpress.mu/article/277175/peut-tout-publier-sur-facebook-cybercrime-unit-demarre-une-campagne
February 2016	Deux internautes arrêtés à la suite d'une plainte de Soodhun.	Two surfers arrested following a complaint made by Soodhun.	http://www.lexpress.mu/article/275235/deux-internautes-arretes-suite-dune-plainte-soodhun
November 2015	Attentats de Paris: L'internaute mauricien presente ses excuses.	Attacks in Paris: the Mauritian surfer apologises.	http://defimedia.info/tag/attentats-a-paris/
October 2015	Cybercrime unit: dans l'univers des enqueteurs du virtuel.	CyberCrime Unit: Investigators in the virtual universe.	http://www.lexpress.mu/article/269976/cybercrime-unit-dans-lunivers-enqueteurs-virtuel
September 2015	Online Racial Hatred Incitement: Police Elaborated a List of 30 suspects.	-	http://news.islandcrisis.net/2015/09/online-racial-hatred-incitement-police-elaborated-a-list-of-30-suspects/
September 2015	Profanation: les auteurs de commentaires violents sur les reseaucx sociaux seront arretes.	Profanation: The authors of violent comments on social networks will be arrested.	http://www.lexpress.mu/article/268334/profanation-auteurs-commentaires-violents-sur-reseaux-sociaux-seront-arretes
August 2015	Video Posted on Facebook: A Man Threatens to disfigure a teenage girl.	-	http://news.islandcrisis.net/2015/08/video-posted-on-facebook-a-man-threatens-to-disfigure-a-teenage-girl/
June 2015	Profil Facebook piraté: Gurib-Fakim portera plainte à la Cybercrime Unit.	Facebook profile hacked: Gurib-Fakim will submit complaint to the Cybercrime Unit.	http://www.lexpress.mu/article/264556/profil-facebook-pirate-gurib-fakim-portera-plainte-cybercrime-unit

May 2015	Chantage: le sexting prend de l'ampleur.	Blackmail: The sexting is growing.	http://www.lexpress.mu/article/261910/chantage-sexting-prend-lampleur
May 2014	Facebook: Yatin Varma porte plainte à la Cybercrime Unit.	Facebook: Yatin Varma complains at the Cybercrime Unit.	http://www.lexpress.mu/article/245869/facebook-yatin-varma-porte-plainte-cybercrime-unit
March 2014	Cyberbullying: Akash Callikan porte plainte à la Cybercrime Unit.	Cyberbullying: Akash Callikan complains at the Cybercrime Unit.	http://www.lexpress.mu/article/cyberbullying-akash-callikan-porte-plainte-la-cybercrime-unit
March 2014	Serie de derapages communaux sur Facebook.	Series of communal skids on Facebook.	http://www.lexpress.mu/article/serie-de-derapages-communaux-sur-facebook
March 2014	Des photos de mineurs en discotheque font polemique sur Facebook.	Minors photos in discotheque controversy on Facebook.	http://www.lexpress.mu/article/des-photos-de-mineurs-en-discotheque-font-polemique-sur-facebook
January 2014	Derive communale sur Facebook: Suzanne Hervet maintenue en cellule policiere.	Communal derivative on Facebook: Suzanne Hervet is maintained in police custody.	http://www.lexpress.mu/article/derive-communale-sur-facebook-suzanne-hervet-maintenue-en-cellule-policiere
July 2012	Propos sectaires sur Facebook: Krishnee Bunwaree bientot face a la justice.	Sectarian comments on Facebook: Krishnee Bunwaree soon facing the justice.	http://www.lexpress.mu/article/propos-sectaires-sur-facebook-krishnee-bunwaree-bientot-face-la-justice