ISSN: 2278-6244
Impact Factor: 6.690

ONLY THUMBS DOWN? SAMSUNG GALAXY NOTE7 FIASCO TWITTER SENTIMENT ANALYSIS USING HYBRID METHOD

Janvier Nkundukozera, Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, South TaiBai Road, Xi'an, China

Ellen Mushimiyimana, Department of Logistics Management, School of Economics and Management, Beijing Jiaotong University, Haidian District Beijing, P. R. China

Roger Nyiringabo, Department of Computer Technology and Engineering,School of computer and information technology, Beijing Jiaotong University, Haidian District Beijing, P. R. China

Evariste Mwumvaneza, Department of Mathematics, Sciences and Physical Education, College of Education and Library sciences, University of Rwanda, Rwamagana-Rwanda

Abstract: *With the almost ubiquitous access on the web, people use microblogs like Twitter, Facebook and Weibo to express their opinions on a wide variety of topics such as products, services, events, organizations, etc. Sentiment analysis on tweets has become a rapid and effective way of gauging opinion for business marketing or social studies. Unlike large opinionated corpora such as products reviews, tweets have unique characteristics that require special treatment to analyze the sentiment they convey. In this paper, we use the hybrid method combining the lexicon based and machine learning-based methods to perform sentiment analysis on tweets in the aftermath of Samsung Galaxy Note7 fiasco. First, we apply the lexicon-based approach todetermine the semantic orientations of opinions expressed in the tweets. This method gives high precision but low recall. To solve this problem, additional opinionated tweets are identified among tweets previously classified as neutral by handling words that are context dependent. We use Pearson's chi-square test to identify opinion words which are not in the lexicon and use them to train a support vector machines (SVMs) classifier to assign polarities to other additional tweets. This method is effective since it does not involve any manual labeling of tweets and has the ability to automatically adapt to new fashions in language, neologisms,and trends found in tweets. In this paper, we found that even though the Galaxy Note7 was removed from the market and killed completely by Samsung due to its battery catching fire, more customers still rated its feature more positive and labeled it as the top Android phone of its time.*

Keywords: *Opinion mining, sentiment analysis, context-dependent opinions, lexicon, SVMs, semantic orientation.*

1. INTRODUCTION

Twitter is a popular microblogging and social networking website on which a big number of people are willing to post their opinions on a wide variety of topics such as products, services, events, organizations, etc. [1]. Thus, Twitter is currently considered as a rapid and effective way of gauging public opinion for business marketing or social studies. Product managers can gain timely insight into opinions on their products by evaluating people's opinions on Twitter[2]. In this paper, we mine people's opinions following the Samsung Galaxy Note7 fiasco.

The Galaxy Note 7 was a flagship phone of Samsung electronics in 2016 and a deemed better android phone of that time[3]. Since its acclaimed launch, the Note 7's early days were marked by glowing reviews because of its amazing features such as a larger, sharper, and richer display than the top phones of that time, less weight , easier to hold, a big phone that didn't feel big, built-in retina scanner, water resistance, rear-facing dual cameras making its specs impressive, the simplicity of its design, the striking beauty of its curved screen and a 3,500 mAH battery that was able enough to allow it to go without a charge even while being used constantly for 36 hours. Nevertheless, it is thought that it is this powerful lightweight battery which includes lithium-ions that would have been its downfall[4]. The early produced Note7 and its replacements did not only got fire in the homes of some customers, shops,and airplanes but also the company themselves decided to halt its production, call customers to exchange Note7 for other Samsung smartphone or receive a refund and finally killed the brand completely after only 53 days of existence from August 19, 2016, to October 10, 2016[5].

In this paper, we gauge the opinions on tweets in the aftermath of the Galaxy Note7 and its features using a combination of lexicon-based and machine learning-based methods to determine the polarities associated to the Note 7 as a whole and to each of its features.

Sentiment analysis can be performed based on different approaches. One of those approaches is based on a function of opinion words in context. Opinion words are words that bear desirable/positive (e.g., good, amazing, etc.) or undesirable/negative (e.g., bad, poor, etc.,) states[6]. This approach uses a dictionary of opinion words to identify and determine sentiment orientation (positive, negative and neutral) in tweets. The dictionary is referred to as the opinion lexicon and the approach of using opinion words to determine opinion orientations is referred to as the lexicon-based approach to sentiment analysis[7][8]. Though this approach has proven itseffective in the analysis of product reviews and can be applied to tweets as well, some tweets characteristics are detrimental to it. For instance, abbreviations (e.g. lol, omg, etc.), emoticons (e.g. :-))) and colloquial expressions (e.g. be blue, go nuts, etc.) etc.which are often used in tweets may convey sentiment orientation but they are not included into the general opinion lexicon.The lexicon-based method would suffer from two problems caused by such forms. First, the lexicon-based method would regard tweets with forms like lol, omg, go nuts, :-) as neutral since these forms are not in a general lexicon. This leads to the law recall problem for the lexicon-based method which depends entirely on the presence of opinion words to determine the semantic orientation of the tweet. It is true that these Twitter expressions can be added to the opinion lexicon but it would almost be in vain since such expressions keep changing and users come up with new ones following the trends and fashions of the internet. Secondly, the polarities of such forms can be domain dependent and this can mislead the lexicon-based method during the calculation of sentiment score. This is the prominent problem associated with the lexicon-based method of sentiment analysis thatrequires a comprehensive lexicon without which the sentiment analysis results will suffer[9].

The alternative to this problem associated withthe lexicon-based method is to apply machine learning-based method to perform sentiment analysis[10]. This is a supervised learning method in which labeled tweets are used to train a classifier that is later used to classify newly acquired tweets into corresponding classes (positive, negative, neutral). However, it is not easy to apply manual labeling in this paper because it would have been labor-intensive and time-consuming to label manually a large set of more than 19500 tweets. To solve all the above-mentioned problems posed by both lexicon-based and machine learning-based methods, a combination of both methods was brought into action and named as the hybrid method. The hybrid method improves both the recall and F-score compared to the lexicon-based method[11][12]. To get the better of the hybrid method, we perform sentiment analysis on tweets as follow. First, we employ a lexicon based-method for tweets sentiment analysis. This methods results in good precision and very low recall. To improve the recall, we apply machine learning-based method on the tweets previously classified as neutral as follows. First, we extract some additional opinionated indicators through the Pearson's chi-square test on the results of the lexicon-based method. Secondly, we train the classifier, support vector machines in our case, to assign sentiment polarities to some of the previously tweets classified as neutral. The hybrid approach is an unsupervised method except for the initial opinion lexicon which is publicly available. The core of the hybrid method of sentiment analysis is the ability to select domain-specific words, emoticons, colloquial expressions, and abbreviations etc. as additional opinion indicators through a statistical test. For example, the positive tweet, "the GalaxyNote7 is so cute. I looooooove it!"Although the expression "looooooove" isnot a general opinion word, if we find itoften co-occurs in positive opinion contexts through a statistical test, we can infer it is a positive opinion indicator. And the SVMs sentiment classifier could learn this important contextual information in training[9].

2. RELATED LITERATURE

This research is in the area of sentiment analysis and to determine whether a tweet expresses a positive or negative sentiment the hybrid method which is a combination of lexicon-based and machine learning-based methods is used.

The lexicon-based approach employs some function of opinion words to determine the polarity or sentiment expressed in the tweet[13][14][7][8]**.** The drawback of this method is that it results in a low recall.

The machine learning-based approach commonly known as supervised learning trains the sentiment classifier using a bag of words features such unigrams or bigrams[10]. The common learning techniques used for sentiment analysis include classification and regression trees (CART)[15], random forests (RF)[16], naïve Bayes (NB)[17], Maximum entropy (Maxent)[18], [19] and support vector machines (SVMs)[20], [21]. All these techniques and others are based on the training data manually labeled following each application domain as it is well known that a sentiment classifier may perform very well in the domain that it is trained, but performs poorly when it is applied to a different domain[22].

There are some literature works that used both the lexicon-based and machine learning-based methods. Classification of reviews into two classes, positive and negative was studied in[23] but the neutral class was left out. A subjectivity lexicon was used to identify training data for supervised learning for subjectivity classification in[24] but our work is not about subjectivity classification. Unlike the hybrid method that identifies additional opinionated tweets among tweets previously classified as neutral, the above literature works did not deal with the third class (neutral class) of the sentiment analysis, hence, they resulted in the low recall[9].

Our work finds the polarity of a tweet as a whole, hence, falls into the document level sentiment analysis like some other works such [25]–[28].

There are also some works related specifically to the tweets sentiment analysis. A classifier to classify tweets into positive, negative and neutral classes was built in [29]. Author of [30] used specific characteristics and language conventions such as hashtags and smiley of Twitter as classifier training features. There are also various online Twitter sentiment analysis systems, for example, Tweetfeel[31], Twendz[32], Sentiment140 [32], etc. All these approaches are based on the supervised learning but the hybrid method used in this paper does not need supervision or manually labeled training data.

3. MATERIALS AND METHODS

This section deals with the exploration of data used in this research to get the glimpse of what is hidden in the tweets under our study. It describes sentiment analysis and the Twitter data characteristics briefly. It also presents the application of the hybrid method on the sentiment analysis of tweets in the aftermath of the Galaxy Note 7 fiasco.

3.1. Characteristics of Twitter data

Tweets are characterized by the Twitter own language conventions[33]–[35]. The following are the example of Twitter conventions.

1. Fixed length. Unlike usually opinionated corpora such as reviews and blogs which could be long, tweets are limited to 140 characters.
2. Emoticons and colloquial expressions are often used in tweets, e.g. looooove, :-).
3. @username. Shows that a tweet is a reply to a user whose Twitter name is "usename1"
4. "#" known as the hashtag is used to mark, organize or filter tweets depending on given topic or category.

5. "RT" an acronym that is put in front of a tweet to indicate that the user is repeating or reposting a tweet.
6. The huge volume of data. It is estimated that 500 million tweets are posted each day. That is 6000 tweets every second and the number is still increasing rapidly[36].

3.2. Sentiment analysis

Sentiment analysis or 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[6][37]. Sentiment analysis mainly studies opinions which express or imply positive or negative sentiments.

The problem of semantic analysis of tweets can be well solved after one has understood properly the structure of an opinion as expressed in tweets.

In general, opinions can be expressed on different things like a product, an individual, an organization, an event, a topic, etc. the term *object* is generally used to denote the entity that has been commented on. An object can be defined as follows[7]:

Object: an object O is an entity which can be a product, person, event, organization, or topic. It is associated with a pair, O: (T, A) where T is a hierarchy or taxonomy of components (or parts), subparts, and so on and A is a set of attributes of O. Each component has its own set of sub-components and attributes.

In practice, both components and attributes are denoted as *features* and this allows to simplify the definition of an object by omitting the hierarchy. In this simplified definition of an object, the object itself is also treated as a feature.

Opinion passage on a feature: the opinion passage on feature *f* of an object evaluated in a tweet *r* is a group of consecutive sentences in *r* that express a positive or negative opinion on *f*.

Opinion holder: the holder of a particular opinion is the person or the organization that holds the opinion. In our case, the tweet holder is the person or the organization that posted the tweet.

Opinion time: opinion time *t* is the time when the opinion is expressed by the opinion holder.

Sentiment or semantic orientation of an opinion: the semantic orientation of an opinion on a feature f states whether the opinion is positive, negative or neutral.

Following the above definitions, a regular opinion can be represented as a quintuple,

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

Where e_i is the name of an entity, a_{ij} is an aspect of e_i, s_{ijkl} is the sentiment on aspect a_{ij} of entity e_i, h_k is the opinion holder, and t_l is the time when the opinion is expressed by h_k. The sentiment s_{ijkl} could be positive, negative, or neutral or expressed with different strength/intensity levels, for example, 1 to 5 stars as used by most reviews sites. The representation of a regular opinion can be explained as follows: the opinion s_{ijkl} must be given by opinion holder h_k about aspect a_{ij} of entity e_i at time t_l.

Even though every tweet we collected has all the components of a regular opinion, in this work we are not interested in opinion holders nor time. This helps to have a model of an object and a set of opinions on the object. This is the basic model used in our paper to study the semantic orientations expressed in tweets. It is known as the simplified model of an opinion and it can be described as follow.

An object is represented by a finite set of features, $F = \{f_1, f_2, \dots, f_n\}$. Each feature f_i in F can be expressed with a finite set of words or phrases W_i, which are synonyms. That is, we have a set of corresponding synonym sets $W = \{W_1, W_2, \dots, W_n\}$ for the n features. Since each feature f_i in F has a name (denoted by f_i), then $f_i \epsilon W_i$. Each author or opinion holder h_k comments on a subset of the feature $S_j \subseteq F$. For each feature $f_k \in S_j$ that an opinion holder h_k comments on, he/she chooses a word or phrases from W_k to describe the feature, and then expresses a positive, negative or neutral opinion on it.

This model introduces three main practical scenarios. Given a collection of tweets D as input, we have:

Scenario 1: Both F and W are unknown. In this scenario, the opinion analysis requires to perform three tasks:

Task one: identifying and extracting object features that have been commented on in each tweet $d \in D$.

Task two: determining whether the opinions on the features are positive, negative or neutral.

Task three: grouping synonyms of features, as different people may use different words to express the same feature.

Scenario 2: F is known but W is unknown. In this scenario, all the three task for scenario one still 1 to be performed, but task three becomes the problem of matching discovered features with the set of given features F.

Scenario 3: W is known, hence, F is known as well. Here we only need to perform task two above which is to determine whether the opinions on the known features are positive, negative or neutral.

This paper deals with scenario 3 since all features of Galaxy Note 7 are well stated in the manufacturer's manual and all authors of tweets use known and standard names to refer to them.

Last but not least, the final output of sentiment analysis should be in an appropriate format. For each tweet evaluated, its output is denoted as a pair (d, SO), where d is a tweet and SO is the semantic or opinion orientation (positive or negative) expressed in d[13]. To present the results, we show the number of tweetsexpress positive or negative opinions on each feature and a graph is also used to give a clear view of opinions on Galaxy Note 7 features expressed in tweets[38].

3.3. Data exploration

In this paper, we use 19,654 tweets collected in aftermath of Samsung Galaxy Note 7 fiasco using GetOldTweets software [39]. We scrapped twitter to gather tweets posted after the launch of Galaxy Note 7 all along to few days after its killing and removal from the market by Samsung. We used the keyword "galaxynote7" and were targeting tweets posted after its launch, its battery catching fire, call for replacements and exchange program, the release of the software patch to limit it charging capacity to go beyond 60 %, flights evacuation, the ban of taking galaxy note 7 for flights, its removal and killing from the market and few days after its fiasco.

To have a glimpse of information hidden in the collected tweets, we utilize the barplot of frequencies of terms, the wordcloud of terms and the barplot showing the number of tweets falling into each of eight National Research Council Canada (NRC) primary emotions.

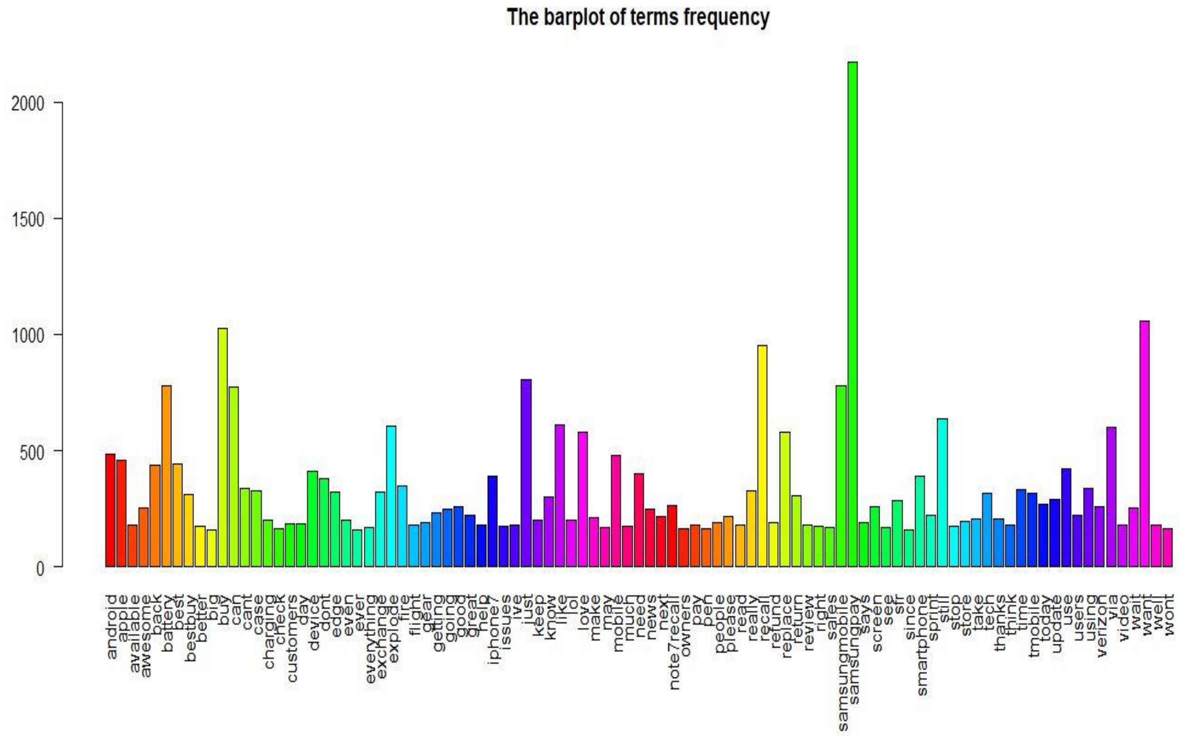

Figure 1: Barplot of terms frequency

On the fig1, we see that terms like Sumsangpay, buy, battery, recall, explode, want etc. have high frequencies are of great significance to our study.

Figure 2: Wordcloud of terms

From the fig2, we can see that terms like Samsungpay, recall, buy, want, battery, replace, explode, love etc. have significant weight and they are the terms occurring often in the tweets about Galaxy Note 7 and significant to our study.

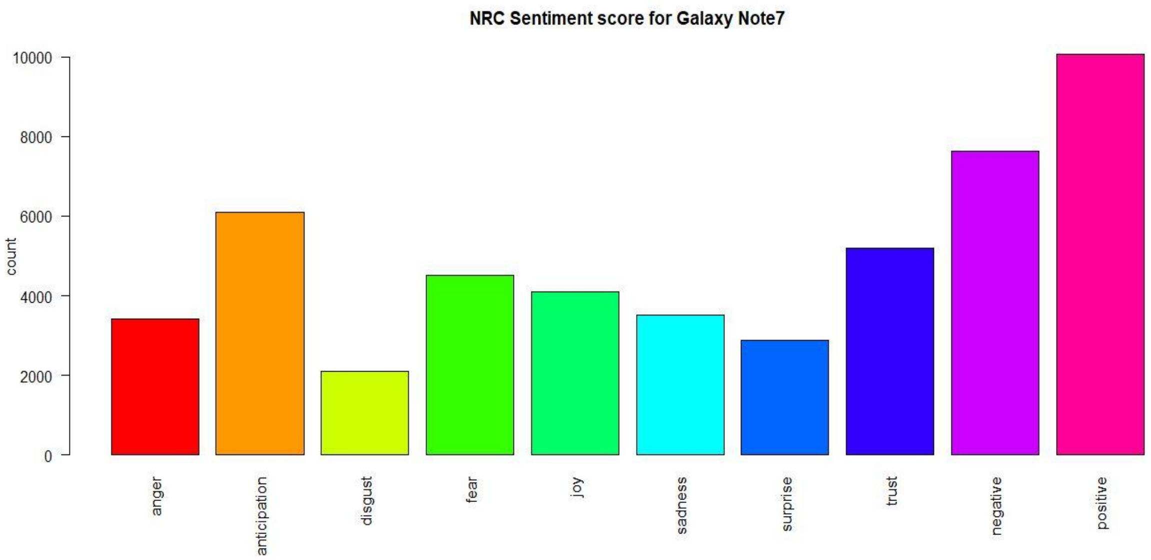

Figure 3: Barplot of the number of tweets corresponding to eachof 8 emotions

Emotion is our subjective feelings and thoughts. Emotions are closely rated to sentiments. The strength of a sentiment or opinion is typically linked to the intensity of certain emotions[6]. According to[40]people have six emotions namely love, joy, surprise, anger, sadness,and fear. On the fig3, emotions such as anticipation, trust, fear, and joy count the big number of tweets respectively. It is also revealed that the number of positive tweets is bigger than negative tweets. This would suggest that more tweets authors expressed their opinion on different features of Galaxy Note 7 with optimism and excitement.

3.4. Sentiment or semantic orientation score calculation

The lexicon-based approach was used to calculate the sentiment orientation score for each tweet. This method uses opinion words. Opinion words are words that encode a desirable state (positive polarity) e.g. love and awesome or undesirable state (negative polarity) e.g. suck and awful. Opinions words can be adjectives, adverbs, verbs,and nouns as well. In this paper, we used a general lexicon from the authors of [13][38]which are two lists of around 6800 positive and negative words. To calculate the sentiment score of every tweet, we used the following algorithm.

Table 1: Calculation of sentiment orientations using the lexicon-based method

#Initialization of variables

Score = 0, Positive_Score=0, Negative_Score=0,

Negation_Score=0

#Match words with the dictionary containing positive sentiment words

If Word== Positive Word,then

Positive_Score = Positive_Score +1

Else

#Match words with the dictionary containing negative sentiment words

If Word==Negative Word, then

Negative_Score = Negative_Score +1

Overall tweet score calculation

Score= Positive_Score - Negative_Score

#decide the polarity of the tweet

If Score>0 Then print

"Positive"

Else

If Score < 0 then print

"Negative"

Else

Print "Neutral"

The lexicon-based method in the table1 resulted inhigh precision since it took care of all opinion words found in tweets and used them to calculate sentiment scores. But it also gave low recall since many opinionated tweets which did not contain opinion words as found in the lexicon were classified as neutral.

3.5. Extraction of opinion indicators

To solve the problem of low recall suffered by the lexicon-based approach, we extracted opinion indicators from tweets previously found opinionated by the lexicon-based method. The indicator would be a word or token which is not in the original opinion lexicon. We use the Pearson's chi-square test to identify those indicators[41]. The theory behind the chi-square test is that if a term is more likely to occur in positive or negative tweets, it more likely to be an opinion indicator. Our task here is to find out how dependent a term *t* is with respect to the positive or negative tweets. Our null hypothesis is that that the candidate indicator *t* is independent of positive/negative tweets with respect to its occurrences in the two sets. The Pearson's chi-square test compares observed frequencies of *t* to its expected frequencies to test the hypothesis.

Table 2: The contingency table for chi-square

	With t	Without t	Row Total
Positive tweets	f_{11}	f_{12}	$f_{11}+f_{12}$
Negative tweets	f_{21}	f_{22}	$f_{21}+f_{22}$
Column total	$f_{11}+f_{21}$	$f_{12}+f_{22}$	

In the table2, f_{ij} represents indicator frequency in the positive/negative tweets. For instance, f_{21} indicates the count of tweets which contain the candidate indicator t in negative tweets.

The chi-square value is computed as follow[9]:

$$X^2(w) = \sum_{i=1,2} \sum_{j=1,2} \frac{(f_{ij} - Eij)^2}{Eij}$$

Where Eij is the expected frequency of f_{21} and is calculated as follows:

$$Eij = \frac{\text{row total}_i \times \text{column total}_j}{f_{11} + f_{12} + f_{21} + f_{22}}, i, j \in \{1, 2\}$$

Finally, we select the indicators with larger chi-square value since the larger the chi-square value, the more dependent *t* is with respect to the positive or negative tweets. We select an opinion indicator if it has a chi-square weight greater than zero.

3.6. Sentiment classifier and hybrid method

In this section, we deal with the last step of the hybrid method which is to train a binary classifier with the newly-indicated opinion indicators. We use Support Vector Machines (SVMs) from the kernlab library [42] as the learning algorithm. Our training data are the newly opinionated indicators and no any opinion indicator originally found in the lexicon of opinion words (positive or negative sentiment words) is used in order to avoid the training bias towards them. That is, the training data consists only of context words.

The hybrid method combines the lexicon-based method and the machine learning-based method to identify the polarity for more opinionated tweets than the lexicon could, hence, the recall is improved. The following is the algorithm to implement the hybrid method:

Table 3: The hybrid method to identify overall opinionated tweets

#Initialization of variables

Score = 0, Positive_Score=0, Negative_Score=0,

Negation_Score=0

#Match words with the dictionary containing positive sentiment words

If Word== Positive Word, then

Positive_Score = Positive_Score +1

Else

#Match words with the dictionary containing negative sentiment words

If Word==Negative Word, then

Negative_Score = Negative_Score +1

Overall tweet score calculation

Score= Positive_Score - Negative_Score

#decide the polarity of the tweet

If Score>0 Then print

"Positive"

Else

If Score < 0 then print

"Negative"

Else

Print "Neutral"

#select additional opinion indicators from opinionated tweets classified by the lexicon-based method.

Opinionated_tweets=positive tweets + negative tweets

Context_words=opinionated_tweets – matched_lexicon_terms

Chi_square_weights= pearsons_chi_square(context_words)

Additional_opinion_indicator=chi_square_weights>0

#Train SVMs classifier and identify additional opinionated tweets

SVM_model=ksvm (Additional_opinion_indicator)

Predictions= predict (neutral_tweets)

Additional_opinionated_tweets=non_neutral_predictions (predictions)

#Overall opinionated tweets

Overall_opinionated_tweets=opinionated_tweets+additional_opinionated_tweets

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this paper, we used the hybrid method which is a combination of the lexicon-based method and machine learning-based method to sentiment classification. The primary goal of the hybrid method is to identify more opinionated tweets, hence, solve the problem of low recall suffered by the lexicon-based method[1], [9], [11], [12], [23], [43]. The results of our experiments show the number of opinionated tweets identified by each method and the number of positive or negative tweets posted about each feature of the Galaxy Note 7.

Table 4: opinionated tweets identified by each method

Sentiment analysis method	**The number of opinionated tweets identified**		
	Positive tweets	Negative tweets	Overall opinionated tweets
Lexicon-based method	5771	4192	9963
Machine learning (additional tweets)	4381	145	4463
Hybrid method	10152	4337	14489

From the table4, we realize that the hybrid method improved the recall of the lexicon-based method dramatically by identifying other more 4463 opinionated tweets. The table4 also shows that 10152 positive tweets and 4337 negative tweets about the Galaxy Note 7 were posted by different authors. That is, a big number of tweet authors were willing to post positive opinions on Galaxy Note 7.

Our experimental results also show the number of positive and negative tweets posted about each of the most prominent features of the Galaxy Note 7[44], [45].

Table 5: The number of positive or negative tweets posted about Galaxy Note7 features

Features	Identified opinionated tweets		
	Positive tweets	Negative tweets	Overall tweets
Galaxy Note 7	**10152**	4337	14489
SamsungPay	**1946**	86	2032
Screen	**179**	42	221
Battery	128	**392**	520
Charging	**179**	48	227
Camera	**122**	29	151
Pen	**134**	14	148
Iris recognition	**95**	10	105
Fingerprint	**5**	3	8
MicroSD card	**14**	0	14
HDR-color	**11**	2	13
Gorilla glass	16	**21**	37
RAM	**18**	4	22
GearVR	**216**	16	232
Exynos chip	**17**	5	22
Marmallow OS	**7**	0	7

Touchwiz	**34**	2	36
GIF creator	**31**	11	42
Securefolder	**11**	0	11
Grace UX	**20**	0	20
Waterproof	**63**	34	97
Size+design+weight	39	**50**	89

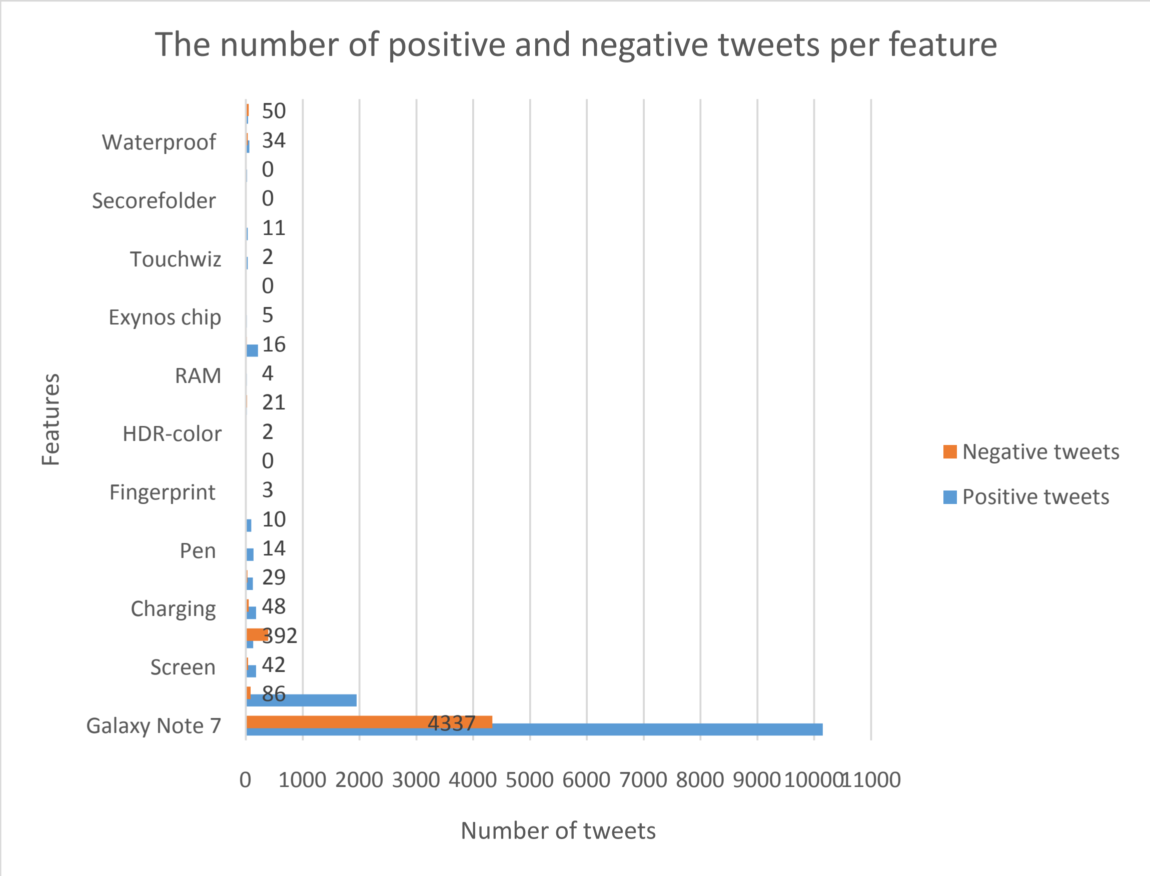

Figure 4: The number of positive and negative tweets per feature

The results in the table5 and the figure4 reveal most of the Samsung Galaxy Note 7 was commented on positively except the battery, gorilla glass, feeling (Size, design,and weight) with 264, 5, 11 negative tweets more compared to positive tweets respectively. The numbers on figure4 show negative tweets since their small numbers can be hard recognized easily following the chat bars.

The figure3, table4 and table5 results all show that more positive opinions were tweeted about the Galaxy Note 7 and its features. That is something stunning considering that this brand of the phonewas removed from the market and killed completely by Samsung. This really shows the potential that was behind this fabulous android phone that pushed many users to resist getting rid of it in exchange to other Samsung phones or money even though it was branded explosive following may of its reports blowing up cars and rooms and flights evacuation following its lithium ions battery catching fire.

5. CONCLUSIONS

With many people able to access the web and willing to post more tweets about different entities like products, events, people, etc. Sentiment analysis on tweets has become a rapid and effective way of gauging opinion for business marketing or social studies.

In this paper, we performed sentiment analysis on the tweets posted in the aftermath of the Samsung Galaxy Note 7 fiasco to find out whether this phone and its features were commented on more negatively or positively by Tweeter users and see whether its removal from the market and killing by Samsung was dictated by the negative opinions expressed by the customers. Hence, the question, only thumbs down? We used the hybrid method which is a combination of lexicon-based method and machine learning-based method to solve the problem of low recall suffered by the lexicon method and find out the answer to our question.

In this paper, we found that the Galaxy Note 7 and its features received more positive than negative opinions, and this justifies the reason why this phone was received with great acclamation and labeled by some customers as the top Android phone of its time. We could argue that the Galaxy Note 7 and its features except its battery were not as bad as some people might think after getting across the story of such a fabulous phone that lived such a very brief and sulfurous life. Hence, not only thumbs down for the Galaxy Note 7.

Acknowledgements: We would like to thank Prof. Jing Liu, for the patient guidance, encouragement and advice she provided for the completion of this paper.

Author Contributions: Janvier conceptualized the paper idea, designed the research design and carried out computations. Ellen, Roger and Evariste carried out data collection, data analysis, and paper writing.

Conflicts of Interest: The author declares no conflict of interest.

REFERENCES

[1] T. K. Lalji and S. N. Deshmukh, "Twitter Sentiment Analysis Using Hybrid Approach," pp. 2887–2890, 2016.

[2] R. Batool, A. M. Khattak, J. Maqbool, and S. Lee, "Precise tweet classification and sentiment analysis," *2013 IEEE/ACIS 12th Int. Conf. Comput. Inf. Sci. ICIS 2013 - Proc.*,

pp. 461–466, 2013.

[3] C. S.-H. Brian X, "Why Samsung Abandoned its Galaxy Note 7 Flagship Phone." [Online]. Available: https://www.nytimes.com/2016/10/12/business/international/samsung-galaxy-note7-terminated.html?_r=0. [Accessed: 08-Jun-2017].

[4] J. Swearingen, "The Brief, Sulfurous Life of the Samsung Galaxy Note 7." [Online]. Available: http://nymag.com/selectall/2016/10/samsung-note-7-killed-and-remo ved -from-market.html. [Accessed: 08-Jun-2017].

[5] S.-H. L. PAUL MOZUR, "Samsung to Recall 2.5 Million Galaxy Note 7s Over Battery Fires," 2016. [Online]. Available: https://www.nytimes.com/2016/09/03/business/samsung-galaxy-note-battery.html. [Accessed: 08-Jun-2017].

[6] B. Liu, *Sentiment Analysis, and Opinion Mining*, no. May. 2012.

[7] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," *Proc. Int. Conf. Web search web data Min. - WSDM '08*, p. 231, 2008.

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.

[9] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," *Hewlett-Packard Dev. Company, L.P*, 2011.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, pp. 79–86, 2002.

[11] G. Vaitheeswaran, "Combining Lexicon and Machine Learning Method to Enhance the Accuracy of Sentiment Analysis on Big Data," vol. 7, no. 1, pp. 306–311, 2016.

[12] S. S. K. Rastogi, "An Improved Sentiment Classification using Lexicon into SVM," vol. 95, no. 1, pp. 37–42, 2014.

[13] M. Hu and B. Liu, "Mining and summarizing customer reviews," *Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '04*, p. 168, 2004.

[14] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," *Proc. 20th Int. Conf. Comput. Linguist. - COLING '04*, p. 1367–es, 2004.

[15] T. Hayes, S. Usami, R. Jacobucci, and J. J. McArdle, "Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two

simulations.," *Psychol. Aging*, vol. 30, no. 4, pp. 911–929, 2015.

[16] L. Breiman, "Random Forests," pp. 1–33, 2001.

[17] D. D. Lewis, "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," *Mach. Learn.*, vol. ECML-98 Te, pp. 4–15, 1998.

[18] A. L. Berger, V. J. D. Pietra, and S. a. D. Pietra, "A maximum entropy approach to natural language processing," *Comput. Linguist.*, vol. 22, no. 1, pp. 39–71, 1996.

[19] N. Mehra, S. Khandelwal, and P. Patel, "Sentiment Identification Using Maximum Entropy Analysis of Movie Reviews," 2002.

[20] Vapnik, "Support Vector Machines," 1979.

[21] T. Joachims, "Introduction to Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Mach. Learn.*, vol. 1398, no. LS-8 Report 23, pp. 137–142, 1998.

[22] A. Aue and M. Gamon, "Customizing Sentiment Classifiers to New Domains: A Case Study.," *Proc. Recent Adv. Nat. Lang. Process.*, vol. 3, no. 3, pp. 16–18, 2005.

[23] S. Tan, Y. Wang, and X. Cheng, "Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples," *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. SIGIR 08*, p. 743, 2008.

[24] J. Wiebe and E. Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts," pp. 486–497, 2005.

[25] M. Gamon, M. Gamon, a. Aue, a. Aue, S. Corston-Oliver, S. Corston-Oliver, E. Ringger, and E. Ringger, "Pulse Mining customer opinions from free text," *Lect. Notes Comput. Sci.*, vol. 3646, pp. 121–132, 2005.

[26] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," pp. 519–528, 2003.

[27] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," *Proc. 2003 Conf. Empir. methods Nat. Lang. Process.* -, vol. 10, pp. 105–112, 2003.

[28] S. R. Das and M. Y. Chen, "Yahoo ! for Amazon : Sentiment Extraction from Small Talk on the Web," pp. 1–16, 2004.

[29] L. Barbosa and J. Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data," *Colin*, no. August, pp. 36–44, 2010.

[30] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced Sentiment Learning Using Twitter

Hashtags and Smileys," *Proc. 23rd Int. Conf. Comput. Linguist. Posters*, no. August, pp. 241–249, 2010.

[31] R. Wauters, "TweetFeel: Real-Time Sentiment Search." [Online]. Available: http://www.washingtonpost.com/wp-dyn/content/article/2009/07/13/AR2009071301691.html. [Accessed: 28-Mar-2018].

[32] J. LOWENSOHN, "PR firm launches Twendz: A Twitter trend analyzer." [Online]. Available: https://www.cnet.com/news/pr-firm-launches-twendz-a-twitter-trend-analyzer/.

[33] C. Greenhow and B. Gleason, "Twitteracy: Tweeting as a New Literacy Practice," *Educ. Forum*, vol. 76, no. 4, pp. 464–478, 2012.

[34] D. Boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2010.

[35] P. Writing, D. H. Challenge, E. Grammar, S. English, and J. Haynes, "Conventions of the English Language in the Twitterverse," pp. 116–125, 2012.

[36] socialpilot, "171 social media marketing statistics for better targetting in 2018," 2018.

[37] B. L. Lei ZhangShuai Wang, "Deep Learning for Sentiment Analysis: A Survey," *WIREs Data Min. Knowl. Discov.*, 2018.

[38] B. Liu and S. M. Street, "Opinion Observer : Analyzing and Comparing Opinions on the Web," *Proc. 14th Int. Conf. World Wide Web*, pp. 342–351, 2005.

[39] Jefferson-Henrique, "GetOldTweets,Copyright (c) 2016 Jefferson Henrique." 2016.

[40] W. G. Parrot, *Emotions in social psychology : essential readings/ed. by W. Gerrod Parrott*. Philadelphia: Philadelphia (Pa.) : Psychology Press, 2001, 2001.

[41] A. Tripathi and S. K. Trivedi, "Sentiment analysis of Indian movie review with various feature selection techniques," *2016 IEEE Int. Conf. Adv. Comput. Appl. ICACA 2016*, pp. 181–185, 2017.

[42] A. Alexandros, A. Smola, K. Hornik, and M. A. Karatzoglou, "Kernel-Based Machine Learning Lab." 2016.

[43] O. Kolchyna, T. T. P. Souza, P. Treleaven, and T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," 2015.

[44] R. P. V. Limited, "Samsung Galaxy Note7." [Online]. Available:

https://gadgets.ndtv.com/samsung-galaxy-note7-3677.

[45] J. WESTENBERGER, “Samsung Galaxy Note 7 officially announced: everything you need to know.” [Online]. Available: https://www.androidauthority.com/samsung-galaxy-note-7-release-date-price-specs-features-702761/ .