Journal of Engineering and Applied Sciences 12 (3): 691-698, 2017
ISSN: 1816-949X

Terrorism Detection Based on Sentiment Analysis Using Machine Learning

Sofea Azrina Azizan and Izzatdin Abdul Aziz
Department of Computer and Information Science, Universiti Teknologi Petronas,
Bandar Seri Iskandar, Perak, Malaysia

Abstract: The advancement in technology especially a micro-blogging site such as Twitter has brought a new era in terrorism where social media is being used as a platform of communication, incite the act of terrorism, recruitment and much more. Terrorist and people supporting this group tend to include sentiment leads to terrorism when sharing their opinions and comments. Thus, sentiment analytics can help to explore and classify the opinion from users to different polarity. Sentiment analysis is an opinion mining process from computer linguistics perspective. There are many existing techniques that have been improved to determine user's opinions in social media but most of the current techniques and algorithms are not explicit to sense the acts of terrorism. Thus, this research is one of the approach to sense user's act leading to terrorism based on the tweets they shared at the Twitter platform. A comparative study between sentiment analysis techniques has been conducted and analysed. In this report, it is proposed to improvise the current sentiment analysis techniques by using machine learning to detect the acts of terrorism more accurately. The novelty about this research is after the sentence have being categorized into positive, negative and neutral categories, all these categories will be compared against the previous sentence of a particular account holder based on the sentiment score for the latest and previous sentence. This means, the tweet's history of a particular account holder on each category will be extracted and the sentiment score calculated. Then, the sentiment score of previous statement will be compared with the sentiment score of the latest statement detected. Machine learning is being proposed to be used in this research as it is more accurate as compared to lexicon-based approach.

Key words: Sentiment, naive bayes, statement, sentiment, accurately

INTRODUCTION

Now a days, social media has been the main platform for people interaction. In most cases, social media users preferred to express their opinions about products, services, events, etc. via the social media rather than using the mainstream media such as TV and newspapers. Numerous data from social media such as blogs, forums, photo-sharing platforms, social gaming and chat applications has contributed towards this trend (Mergel, 2013).

According to Twitter Inc., there are 320 mln. of monthly active users around the world engaged with Twitter (2016). Approximately there are 6000 tweets every second, resulted in approximately 350,000 tweets being consistently sent by users each minute. Thus, there are over 500 mln. tweets been generated by all users around the world per day. Twitter has reached >465 mln. accounts in 2012 as reported in the news. The users are tweeting on numerous topics, comments, opinions and thoughts regardless it is good or bad.

Some users are abusing the capability of this social media to spread distorted beliefs and negative influence to other users. These include terrorism, politics, religions, fraudsters, ideology and others. In 2015, >125,000 user accounts are linked to terrorists and have been deleted by Twitter (Yadron, 2016). An in famous terrorist group named Islamic State in Iraq and Syria (ISIS) uses social media like Twitter as the main communication platform to strategize and to gain supporters via hashtags statements or phrases (Concordiam, 2014).

Other than that, Fajr-al-Bashaer (Dawn of Good Tidings) has been developed by ISIS and made available in Google Play where it can send the information such as to recruit, radicalise and raise funds and updates from Syria and Iraq directly to the user in 2014 (Berger and Morgan, 2015). Moreover, ISIS uses Twitter for recruiting, to spread ideology and to give explicit instructions to specifically target people such as extremist group with jihadist mentality, youth who supports terrorism and others (Younas, 2014).

Corresponding Author: Sofea Azrina Azizan, Department of Computer and Information Science, Universiti Teknologi Petronas, Bandar Seri Iskandar, Perak, Malaysia

Tweets leading to terrorism are not being detected accurately by current sentiment analysis techniques because sometimes the statements involve jokes, sarcasm or maybe a real threat. In general, sentiment analysis techniques are used to determine the opinion regarding a specific issue. The current sentiment analysis is lacking of technique detecting on terrorism. It could not capture the right parameter such as keywords, user behavioural analytics and tweets pattern to indicate the sentiment of terrorism.

All of the above issues show that Twitter is being used as a platform to in-doctrine terrorism among the social media users. With the number of accounts and the amount of tweets increasing exponentially, the number of terrorism influencing activities also increased. Thus, the sentiments implicit in Tweets related to terrorism is difficult to interpret.

Problem statement: The existing methods such as support vector machine (Medhat *et al*., 2014), neural network (Santos and Gatti, 2014) and naive bayes are still lacking the ability to interpret the true intention of the Twitter account holder. There are still lacking of techniques to detect sentiment on terrorism in Twitter. The current sentiment analysis methods are still inferior because it does not include user behavioural analytics and specific parameters.

Such specific parameters are keyword related to terrorism, tweets history, tweets pattern and user profile. To my concern, currently there is no viable techniques yet that can analysis, sense, detect and filter tweets specially related to terrorism. Most of the techniques classifying the sentiment based on one sentence tweeted from a particular user only. This resulted to less accuracy because all statements involved jokes, sarcasm and may be a real threat which make it difficult to classify the sentence accurately. Logically, by adding a certain parameter the accuracy can be increased. The user behavioural analytics must be added on in the sentiment classification because we need to study the tweets pattern of a particular holder to enhance the classification result.

Research questions:

- Is the existing sentiment analysis techniques able to estimate the sentiment expressed in tweets about terrorism based on specific parameters leading to terrorism?
- How can tweets that relate to terrorism be analysed using the machine learning techniques?
- To what extent of analytical accuracy can the machine learning techniques achieve in analysing sentiments in tweets that leads to terrorism?

Research objectives:

- To investigate the existing sentiment analysis's techniques used to analyse sentiments expressed in tweets about terrorism
- To propose a method that can sense, filter and analyse sentiments leading to terrorism in tweets using ML techniques
- To validate the proposed technique's accuracy in sensing, filtering and analysing sentiments leading to terrorism

Literature review

Terrorism in social media: Social media and terrorism are very synonymous nowadays. Social media applications such as Twitter, Facebook, Instagram, Snapchat and many others are actively being used by terrorist to manifest terrorism in the global arena. They use social media as the main medium to inject the idea of jihad and capture people attention about their activities. ISIS is now aggressively embracing the social networks and to intimidate people and inspire them about the false jihad. Based on Recorded Future's Index Website the increasing number of references Twitter about ISIS is growing higher as compared to other social media, since 1st January 2014. The number of tweets generated about ISIS from 1st January 2014 until 1st February 2015 as compared to other media. The highest tweets been generated was on 1st February 2015 which amounting to >1,200,000 tweets.

In Malaysia, the terrorism issue is deemed expanding with the increasing number of people being arrested as related to ISIS. According to Hata, he reported that there are 15 people suspected to relate to ISIS extremist group and planning a terror attack in Malaysia (Wahari, 2016). Reported from the others, those 15 arrested people are also gathering funds for a militant group, hacking into several government-owned websites, recruiting Malaysians and sending them to Syria for battle (Ramendran, 2016).

Thus, a technique called sentiment analysis or opinion mining is being used to identify and study about user behaviour through the information extracted from social media. Therefore, there is a need to propose a new improvised algorithm which can categorize the opinions in social media to predict the user interest or behaviour especially towards the terrorism sentiment.

Sentiment Analysis (SA): Sentiment Analysis (SA) is becoming one of the attractive research topic in computer science (Feldman, 2013). Sentiment analysis is known as opinion mining which is a computational study used to gather the overall attitude toward some specific item based on opinions (Liu, 2010). Sentiment analysis is treated as a classification task as it classifies the orientation of a text into either positive or negative and neutral. Using opinion mining, tweets are extracted and then the text is parsed and compared to a list of keywords to categorize them into classes.

In general, sentiment analysis is concerning with two types of textual information which in example facts and opinions. Facts are objective expressions about the events and entities while opinions are subjective that can describe people sentiments, appraisal or feelings towards some entities. So far, the current research in sentiment analysis mainly focused on two things which are identifying class or it is called a sentiment polarity of the texts and identifying a given text whether the semantic is subjective or objective (Liu, 2010).

Based on Fig. 1, the two main approaches in sentiment analysis are the machine learning and the lexicon-based approach. Machine learning will be further elaborated while lexicon-based approach. Both of the approaches are being classified between supervised learning and unsupervised learning. Supervised learning means the input which is the datasets are given and we know what is the expected output. In unsupervised learning, the datasets are given but we do not know what would be the expected output.

Based on Table 1, training data for supervised learning must be labelled while the unsupervised data are not labelled. Labelled data means the data is consisting of a label of input and output. This is the reason why supervised learning knowing the input of the data and expected output as mentioned above. With labelled data, the period for supervised learning to train the data is shorter. Furthermore, supervised learning is more accurate as compared to unsupervised learning. The accuracy in sentiment analysis in this context is referring to a measure of how often a sentiment rating was correct. Sentiment analysis suffers inaccuracy because there are flaws such as sentiment ambiguity, sarcasm, positive and negative words in the sentence that are misinterpreted.

Machine learning approach: Machine learning methods are also known as the supervised approach because it is frequently relying on supervised classification learning (Goncalves *et al.*, 2013). Sentiment detection in machine learning is trained to determine positive, negative and neutral sentiments classification. Machine learning consists of various types of technique which include support vector machine, neural network, naive bayes, bayesian network and max entropy.

From Table 2, naive Bayes is chosen to be used in the future phase. Even though, the accuracy of naive bayes is slightly lower as compared to SVM, it is found to suit

Table 1: Difference between supervised and unsupervised learning

Characteristics	Supervised	Unsupervised
Data to train classifier	Labelled	Unlabelled
Period to train the data	Shorter	Longer
Accuracy	High	Low
Advantages	Models that can easily be understood	Can detect unknown data
Disadvantages	Cannot detect unknown data	Unpredictable result

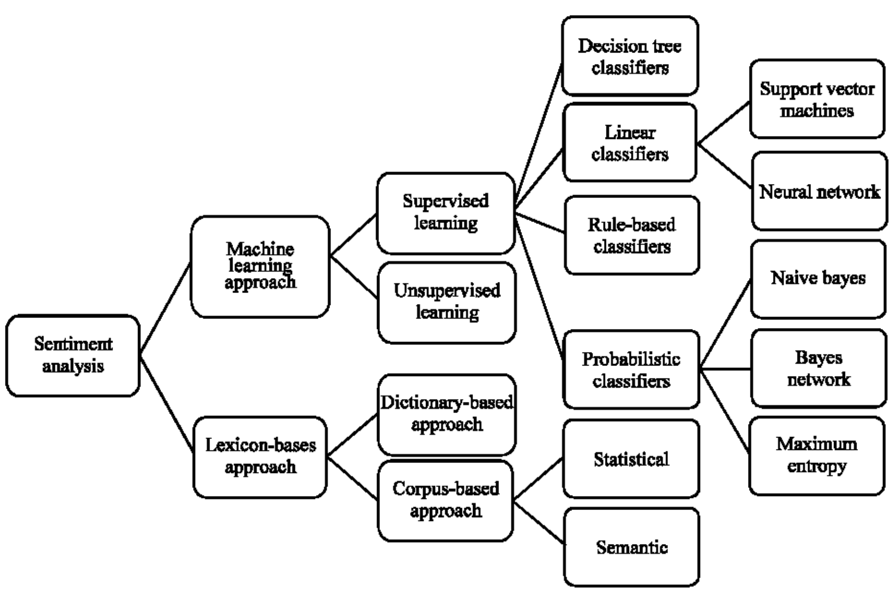

Fig. 1: Sentiment classification techniques used in sentiment analysis (Medhat *et al.*, 2014)

the research better because it offers advantages in terms of less training time needed as compared with SVM and neural network. Furthermore, naive Bayes is highly scalable and in particularly useful for very large data sets like Twitter data (Ray, 2015). However, naïve Bayes suffers low accuracy if using less training data.

Figure 2 shows a basic step in machine learning approach. In general, firstly, the data from social media to be analysed is gathered. In this research, the data will be collected through Twitter streaming API that has been used by Zhang in his study. The data will be extracted based on the user criteria, for example by matching a keyword of "terrorism". Upon the completion of the data gathering phase, the data will undergo a pre-processing phase. In this phase, the data is cleaned by filtering and removing all URLs, @tags, #hashtags, spelling error, non-english words, correcting the lowercase and uppercase and many more. Next, the data are being prepared and tagged with the input and output label. Then, the data is classified into sentiment polarity which are positive, negative or neutral class by using any machine learning method.

Lexicon based approach: Lexicon based approach makes the use of dictionaries only without any training data to train the classifier. There are two types of techniques in lexicon based which are dictionary-based method and corpus-based method. Lexicon-based approach is using a dictionary which are wordnet, sentiwordnet and q-word as being used by other researchers to match the words from the statement (Medhat *et al*., 2014; Khan *et al*., 2015). Based on the dictionary, the scoring words have their own weightage. Thus, the classification can depend on the overall scoring sentence or only by words. The statement will be compared with database words and once there are newly words found, it will be added into the list of words. Then, the iteration will stop until there are no new words found in the sentence. Corpus-based approach is focused on text context rather than words. It takes advantage of a syntactic pattern which means it based on the syntax or rules to identify new sentiment words and their polarity in a written text or it called corpus. The objective or corpus based is to find another opinion words in a large corpus by scanning through phrases that match certain part of another sentence.

Based on Table 3, all researchers from different studies concluded that machine learning gives a better result as compared to lexicon-based approach. According to the studies, all researchers agreed that machine learning gives a high accuracy of classification result. Based on Hailong *et al*. (2014), the disadvantage of machine learning is the result can be biased. It means, since the data of most machine learning technique is supervised (which the data are labelled) therefore, the input and expected output are known. However, lexical-based does not need labelled data since it is unsupervised learning. According to Pavel machine learning is taking a shorter period to train the data but it cannot detect an unknown data like the lexical-based. After having the comparison between these two main techniques, this research is proposed to adopt a machine learning approach.

Table 2: Comparative study between machine learning techniques (Medhat *et al*., 2014; Feldman, 2013; Thakkar and Patel, 2015)

Techniques/ parameter	Support Vector Machine (SVM)	Neural network	Naive bayes
Accuracy	High	Low	Medium
Training time	High	High	Low
Advantages	Can better cope in many noisy features; Hard to interpret the data	Easy to use with few parameters	High scalability of data; fast classification rate
Disadvantages	Long training time needed	Long training time needed	Low accuracy if using less training data

Table 3: Comparative study between machine learning and lexicon-based from other researchers (Thakkar and Patel, 2015)

Researcher	Machine learning	Lexical-based	Opinion
Blinov *et al*. (2013)	Advantage: high accuracy of classification result Disadvantage: need of labels data	Advantage: no need of labelled data Disadvantage: dictionary is required	Machine learning demonstrated good result as compared to lexicon based approach
Zhang	Advantage: give higher precision Disadvantage: the result may fail if the data are biased when training	Advantage: requires strong linguistic resources (dictionary) Disadvantage: not adapt well in difference domains or languages	Machine learning such as SVM and naive bayes give better result as compared to lexicon based
Harsh	Advantage: high accuracy and adaptability Disadvantage: Need a well design classifier	Advantage: does not require data training Disadvantage: low accuracy when the data size is high	Machine learning delivered a better result as compare to lexicon However, it is depending on the datasets

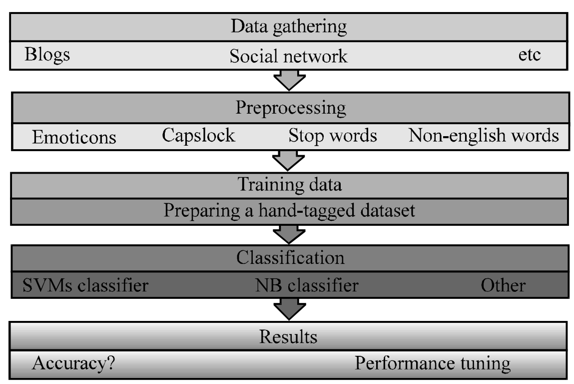

Fig. 2: Machine learning steps (Ray, 2015)

MATERIALS AND METHODS

System development: The purpose to have a research methodology diagram is to give the insights about how the thought of this research is delineated into chart to give an outline and thought to the reader (Fig. 3).

Data gathering: In this research, the micro-blogging service which is Twitter is chosen to be the data source because it is the main platform for terrorism communication nowadays. Twitter has recorded the highest issues related to the acts of terrorism as compared to Facebook and Blogs. The data are collected from Twitter's streaming application programming interface or API. The user can request type of the data they intend to extract by setting a set of criteria such as keywords, usernames, locations, name place and so on. Other than that, with the Twitter API, real time tweets can be extracted as well. In this research, the data criteria are based on terrorism keywords such as "ISIS", "Muslim", "bomb", "terrorists" and etc. As the tweets match with the user's criteria, they are pushed directly to the user in json format. Json format is a javascript object notation which it is an open standard format of a document.

Data pruning: After data is gathered, data pruning process is deployed. Data pruning or pre-processing is needed to normalize the data. Example of data pruning are removing URL, @tags hashtags, uppercase and lowercase to common case, spelling error and etc.

Mapping: In mapping, sentiwordnet is used. Sentiwordnet consists of thousands of english words which have been attributed to a positive or negative score. Thus, based on the database word, tweets sentence will be compared and calculate the score by referring to sentiwordnet dictionary. Since, the word alone is not really enough to make a decision, thus the total score will be calculated based on the context of the sentence or the phrase as well.

Sentiment classification: Sentiment classification also known as sentiment orientation is to categorize the sentence into the classes of positive, negative or neutral. Naive bayes is being proposed because it is mostly used by researcher in sentiment analysis. Thus, many references can easily be accessed during a development phase. It uses Bayes theorem to predict the probability that a given feature set belongs to a particular label. With the equation given, the naive bayes will classify the statement to positive, negative or neutral based on the sentiment score result. Discussion on naive bayes has been addressed.

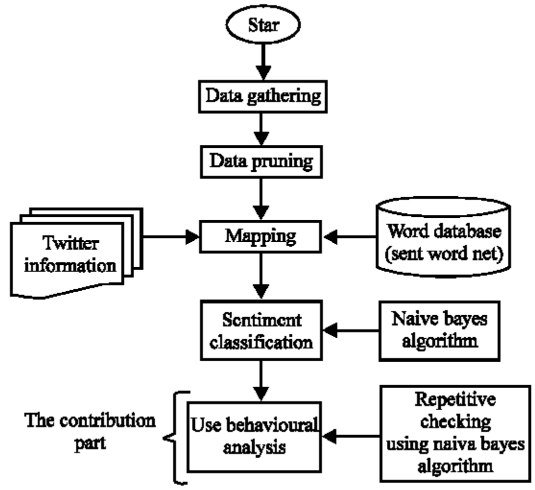

Fig. 3: Proposed system development diagram

User behavioural analytics: At this phase, snapbird tool is used to track down a particular user tweet's history according to the specific tweet handle or name of Twitter's account holder. As soon as the tweets has been classified to their specific polarity based on the sentiment score of phrases, all the three classes will be checked repetitively. In order to undergo the second checking, tweets on each category will be compared with their tweet's history. The purpose of repetitive checking is to get a better accuracy result when determining the tweets whether it is leading towards a terrorism or not. Accuracy means, we will check the previous tweets sentiment score and compare it with the current tweets.

Figure 4 depicts the process flow in user behavioural analysis. If the sentiment score is classified as negative, then after the repetitive checking and resulted on the same class, it can be concluded that this account holder is leading towards the acts of terrorism. For example, if the user tweeted "Muslims are terrorist" and the sentiment score is classified as negative class, then this particular tweet will be compared to the specific user's tweets history.

The purpose to check user's previous tweets is to analyse their tweets patterns. Tweets patterns are being referred as the style of tweets from a particular user. Sometimes, the user themselves is someone who loves to joke about terrorism then tweet patterns can be categorized as jokes. There is also a user who is serious to talk about terrorism with an intention to support and wanted to influence other reader about terrorism.

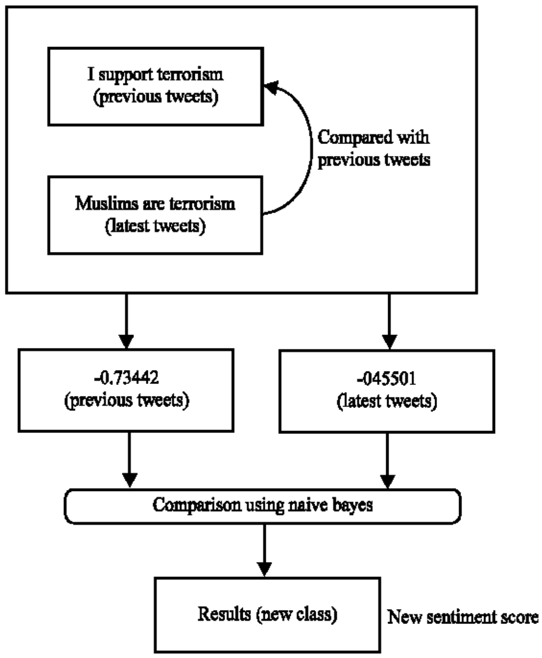

Fig. 4: Block diagram of user behavioural analysis

RESULTS AND DISCUSSION

Proof of concept

Critical analysis naive bayes from other research studies: In this study, a proposed sentiment analysis technique is being proposed. This research proposes to use naive bayes approach after a comparative analysis in Chapter 2. A comparative study of naive bayes between different researchers is shown in Table 4.

In view of the three studies that use the naive bayes technique in sentiment analysis research, this technique has been proven and has a potential to be implemented in this study. Based on these studies, the researchers mentioned that naive bayes performance is fast and simple to be applied. The accuracy of naive bayes is recorded between 85-96%. Other than that, the researchers concluded and suggested before any researcher would like to apply this technique, considering to have the right type of feature selection of the information first. This purpose is to increase the performance and accuracy resulted at the end of the research. As mentioned by Hassan *et al.* (2012), a quality data will affect the performance of naive bayes technique.

Naive bayes algorithm: As stated earlier, the naive bayes technique is being proposed in this research. The Baye's theorem is being applied to predict a class for any given text from the tweets. Below is the formula to be adopted (Toit, 2015):

$$P(\text{label}|\text{features}) = \frac{P(\text{label})P(\text{features}|\text{label})}{P(\text{features})} \quad (1)$$

From the formula, P (label) can be read as the class of the tweets whether it is positive, negative or neutral while P (features) is the tweets. P (label|features) is the end result of this technique. Below is the example of tweet and its class:

$$P(\text{positive}|\text{tweet}) = \frac{P(\text{positive})P(\text{tweet}|\text{positive})}{P(\text{tweet})} \quad (2)$$

The main part is since the P(tweet) is constant. Then, we can substitute the formula with the weightage value of the sentence. In general, naive bayes has been proven as an easy implementation in sentiment analysis area. P(positive|tweet) means the probability of a particular tweet being labelled as positive. P(tweet|positive) means the times of the words in a particular tweet has appeared in tweets labelled as positive in training datasets.

P(Positive) means the possibility of the tweet's class. For example, there are three possible classes which are positive, negative and neutral. It gives any tweet a one in three (33%) chance of falling into any of those classes. It will result to P (positive) = 0.33333.

P(tweet|positive) needs of a training set of tweets that were already classified into the three categories. This can be referred to the word database which is sentiword.net. Then, tokenize the tweet and calculate the probability for each word in the training set. Tokenization means the sentence is separated one by one. This the formula to calculate word by word:

$$P(\text{tweet}|\text{positive}) = P(\text{T1}|\text{positive})*P(\text{T2}|\text{positive})*...*P(\text{Tn}|\text{positive})$$

T1, T2 until Tn are all the words in the tweets. For P(T1|positive), it is to determine the probability of a specific word falling into the category we are testing. The total number of times for a particular word occurs in a positive tweet has to be divided with the total number of words in positive tweets. Both of these are taken from the training datasets list. This calculation of words need to be repeated for each word in the tweet.

Table 4: Comparison of naive bayes technique from other studies (Narayanan *et al*., 2013; Xhemali *et al*., 2009; Hassan *et al*., 2012)

Reserarchers	Study title	Domain	Performance	Accuracy (%)	Conclusion and future work
Vivek	Fast and accurate sentiment classification Using an enhanced naive bayes model	Movie review	Extremely fast to train the data	88.0	Choosing the right type of feature and removing of noises from the data
Daniela	Naive bayes vs. decision trees vs. neural networks in the classification of training web pages	Web pages data	Fast as it needs little training time; fast in predicting the class	95.2	Increase the parameters to not only focus on the content but see also the title, link information and meta data
Hassan *et al*. (2012)	Alleviating data sparsity for twitter sentiment analysis	Stanford Twitter Sentiment dataset	Simple and straight-forward and comparable with other techniques	86.3	Selective during feature selection. A quality data affects the performance and should attach a weight for each extracted sentiment topic

After all, the same procedures for the other two classes (negative, neutral) is repeated. Lastly, all of the tree probabilities calculated for all classes are compared. The highest ranked class is chosen as the class for the document.

Data preparation: The preliminary result of this research is through extracting and analysing tweets based on top terrorism keywords listed in the United States Department of Homeland Security (DHS) for social media (2011). Based on the DHS, there are >50 words listed as terrorism keywords such as (terrorism, jihad, bomb, radical, Al-Qaeda, Abu Sayyaf, isis, extremist, Hezbollah, suicide bomber, chemical weapon and many more). Based on the terrorism list of keywords, top eight words from the list that are popularly being used are extracted. The words are (terrorism, jihad, bomb, radical, Al Qaeda, Abu Sayyaf, isis and extremist).

Data is extracted through followthehashtags.com which is an intelligence tool to analyse Twitter data. The objective of extracting the words as a preliminary result is to analyse whether Twitter contains the data based on the keywords requested. Other than that, it gives an early insight of what types of data retrieved to be used in the development phase later.

Based on Table 5, the keywords extracted from the operators with “OR” separator between each words. For example, terrorism OR jihad OR bomb OR radical OR al qaeda OR abu sayyaf OR isis OR extremist. In referring to Table 5, the total tweets are extracted between one day which is from 30-31 March 2016. From the result, the total audience or users gained within a day are >20 mln. but the number of contributor on related keywords are 1334. From the number of contributors, there are 1480 tweets generated which is higher from the number of contributors. This can be said that, there is one contributor who tweets more than once on the related keywords. As been recorded in the table, tweets per contributor ratio is 1:11 which means one person contributes one or more tweets. The total impressions are 24,923,206 recorded within a day. Impressions are like an impact factor in Twitter. It simply means the number of users reached by the tweets to other users.

Table 5: Summary of tweets based on requested keywords

Key word	Terrorism, jihad, bomb, radical, Al-Qaeda, Abu Sayyaf, isis, extremist
Total audience	20,133,060
Contributors	1,334
Total tweets	1,480
Total potential impressions	24,923,206
Measured data from	2016-03-30 13:10
Measured data to	2016-03-31 11:14
Tweets per contributor	1:11
Impressions/audience	1:24
Measured time in seconds	79,403
Measured time in minutes	1,323
Measured time in hours	22
Measured time in days	1
Tweets per second	0.01863
Tweets per minute	1.11834
Tweets per hour	67.1007
Tweets per day	1610.417

CONCLUSION

Based on the critical analysis, naive bayes is chosen as the technique to be adopted and improved in this research. The naive bayes technique is recorded as a medium accuracy as compared to support vector machine and neural network. We have discussed how to increase the accuracy in sentiment classification by using naive bayes.

The novelty of this research is to improve the algorithm of naive bayes on detecting a sentiment that leads to terrorism on Twitter. In order to increase the accuracy, the element of user behavioural analysis has been proposed to embed into the algorithm after the sentiment classification process have been done. The user behavioural analytic will be focused on the retrieval of tweet’s history from a particular user based on specific keywords related to the terrorism issue. Then, the latest and previous tweets of a particular user’s sentiment score will be compared. A new class of sentiment (positive, negative and neutral) is being used based on the comparison of the latest and previous tweets.

REFERENCES

Berger, J.M. and J. Morgan, 2015. The ISIS Twitter census defining and describing the population. Brookings Project U.S. Relat. Islamic World, 3: 1-65.

Blinov, P., M. Klekovkina, E. Kotelnikov and O. Pestov, 2013. Research of lexical approach and machine learning methods for sentiment analysis. Comput. Linguistics Intellectual Technol., 2: 48-58.

Concordiam, P., 2014. Extremism hits home stopping the spread of terrorism. J. Eur. Secur. Defense Issues, 5: 1-67.

Feldman, R., 2013. Techniques and applications for sentiment analysis. Commun. ACM., 56: 82-89.

Goncalves, P., M. Araujo, F. Benevenuto and M. Cha, 2013. Comparing and combining sentiment analysis methods. Proceedings of the 1st ACM Conference on Online Social Networks, October 07-08, 2013, ACM, Boston, Massachusetts, ISBN: 978-1-4503-2084-9, pp: 27-38.

Hailong, Z., G. Wenyan and J. Bo, 2014. Machine learning and lexicon based methods for sentiment classification: A survey. Proceedings of the 11th Conference on Web Information System and Application Conference (WISA), September 12-14, 2014, IEEE, Nanjing, China, ISBN:978-1-4799-5726-2, pp: 262-265.

Hassan, S., H. Yulan and A. Harith, 2012. Alleviating data sparsity for Twitter sentiment analysis. Proceedings of the 2nd Workshop on Making Sense of Microposts: Big Things Come in Small Packages and 21st International Conference on World Wide Web, April 16, 2012, CEUR, Lyon, France, pp: 2-9.

Khan, A.Z., M. Atique and V.M. Thakare, 2015. Combining lexicon-based and learning-based methods for Twitter sentiment analysis. Int. J. Electron. Commun. Soft Comput. Sci. Eng., 2015: 89-91.

Liu, B., 2010. Sentiment Analysis and Subjectivity. In: To Appear in Handbook of Natural Language Processing, Indurkhya, N. and F.J. Damerau (Eds.). 2nd Edn., University of Illinois, Chicago, USA., pp: 1-38.

Medhat, W., Hassan, A. and H. Korashy, 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J., 5: 1093-1113.

Mergel, I., 2013. Social media adoption and resulting tactics in the US federal government. Government Inf. Q., 30: 123-130.

Narayanan, V., I. Arora and A. Bhatia, 2013. Fast and accurate sentiment classification using an enhanced Naive Bayes model. Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, October 20-23, 2013, Springer, Berlin, Germany, pp: 194-201.

Ramendran, C., 2016. Cop among 15 detained for IS militancy. The Sun Daily, Malaysia. http://www.thesundaily.my/news/1739087

Ray, S., 2015. 6 East steps to learn naive bayes algorithm. Analytics Vidhya, USA. https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/

Santos, C.N.D. and M. Gatti, 2014. Deep convolutional neural networks for sentiment analysis of short texts. Proceedings of the COLING 25th International Conference on Computational Linguistics, August 23-29, 2014, Dublin City University, Dublin, Ireland, pp: 69-78.

Thakkar, H. and D. Patel, 2015. Approaches for sentiment analysis on Twitter: A state-of-art study. Department of Computer Engineering, Montreal, Quebec, Canada.

Toit, J.D., 2015. The Bayes classifier: Building a tweet sentiment analysis tool. Cloud Academy Blog, USA. http://cloudacademy.com/blog/naive-bayes-classifier/

Wahari, H., 2016. Malaysia arrests 15 alleged IS supporters. Benar News, Kuala Lumpur, Malaysia. http://www.benarnews.org/english/news/malaysian/suspects-arrest-03242016144757.html

Xhemali, D., C.J. Hinde and R.G. Stone, 2009. Naive Bayes vs. Decision trees vs. Neural networks in the classification of training web pages. Intl. J. Comput. Sci. Issues, 4: 16-23.

Yadron, D., 2016. Twitter deletes 125,000 Isis accounts and expands anti-terror teams. The Guardian, London, England, UK.

Younas, M.A., 2014. Digital jihad and its significance to counterterrorism. Counter Terrorist Trends Anal., 6: 10-17.