IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 January 2023, accepted 9 February 2023, date of publication 17 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3246162

RESEARCH ARTICLE

Political Security Threat Prediction Framework Using Hybrid Lexicon-Based Approach and Machine Learning Technique

NOOR AFIZA MAT RAZALI[1], NUR ATIQAH MALIZAN[1], NOR ASIAKIN HASBULLAH[1], MUSLIHAH WOOK[1], NORULZAHRAH MOHD ZAINUDDIN[1], KHAIRUL KHALIL ISHAK[2], SUZAIMAH RAMLI[1], AND SAZALI SUKARDI[3]

[1]Faculty of Defence Science and Technology, National Defence University of Malaysia, Kuala Lumpur 57000, Malaysia
[2]Master of Technology, Management and Science University, Selangor 40100, Malaysia
[3]CyberSecurity Malaysia, Selangor 63000, Malaysia

Corresponding author: Noor Afiza Mat Razali (noorafiza@upnm.edu.my)

This work was supported in part by the National Defence University of Malaysia (UPNM) under Grant PS0031–UPNM/2022/GPPP/SG/9, and in part by the Fundamental Research under Grant FRGS/1/2021/ICT07/UPNM/02/1.

ABSTRACT The internet offers a powerful medium for expressing opinions, emotions and ideas, using online platforms supported by smartphone usage and high internet penetration. Most internet posts are textual based and can include people's emotional feelings for a particular moment or sentiment. Monitoring online sentiments or opinions is important for detecting any excessive emotions triggered by citizens which can lead to unintended consequences and threats to national security. Riots and civil war, for instance, must be addressed due to the risk of jeopardizing social stability and political security, which are crucial elements of national security. Mining opinions according to the national security domain is a relevant research topic that must be enhanced. Mechanisms and techniques that can mine opinions in the aspect of political security require significant improvements to obtain optimum results. Researchers have noted that there is a strong relationship between emotion, sentiment and political security threats. This study proposes a new theoretical framework for predicting political security threats using a hybrid technique: the combination of lexicon-based approach and machine learning in cyberspace. In the proposed framework, Decision Tree, Naive Bayes, and Support Vector Machine have been deployed as threat classifiers. To validate our proposed framework, an experimental analysis is accomplished. The performance of each technique used in the experiments is reported. In this study, our proposed framework reveals that the hybrid Lexicon-based approach with the Decision Tree classifier recorded the highest performance score for predicting political security threats. These findings offer valuable insight to ongoing research on opinion mining in predicting threats based on the political security domain.

INDEX TERMS Cyberspace, lexicon-based approach, machine learning, national security, opinion mining, political security, sentiment analysis.

I. INTRODUCTION

Cyberspace has become an important paradigm in the national security domain. According to the Worldwide Threat Assessment of the US Intelligence Community (2016), cyber-related threats are among the prominent threats in line with terrorism, the proliferation of weapons of mass destruction and counter intelligence [1]. Securing a country is more complicated in modern times compared to previous decades. In this era, big data, massive information, online rumours and fake news are constantly shared in cyberspace. This can evoke negative emotions and disruptive behaviour, which may jeopardise national security.

The associate editor coordinating the review of this manuscript and approving it for publication was Chong Leong Gan.

Researchers have found that a strong relationship exists between opinions or sentiments triggered by emotions and national security threats. It was further noted that sentiments, also known as opinions, included in a text can provoke negative feelings or elicit emotions such as rage or fear which can trigger events that threaten national security. Since information shared in cyberspace is frequently embedded with emotions that may contain national security threats (according to each element of national security), real-time detection of disruptive emotions plays a key role in helping authorities manage the situation early. Various gaps, techniques and domain applications that focus on existing opinion mining methods (such as the lexicon-based approach and machine learning techniques) can be used to determine the existing sentiments embedded in sentences throughout several domains, as discussed in [2].

The assessment and framework analysis regarding emotions and their measurements in the aspect of national security are lacking. Opinion mining-related research in the national security domain has not been fully explored, although it can determine various threats and aid in the protection of a nation. Thus, this area requires comprehensive research [3]. Previous studies have mainly focused on how human emotions can be classified using various methods. Less attention has been given to the relationship between emotions and national security threats as well as the methods to predict whether or not such threats are increasing. In this study, we propose a new theoretical framework for predicting political threats which we suggest are highly related to emotions embedded within the text of online news. The scope of this research is political security which is a key element of national security. The proposed framework is validated by experimental analysis using the hybrid technique in mining people's sentiments or opinions, which also includes the emotional aspect of political security. This is accomplished using a combination of the lexicon-based approach and machine learning techniques which are Decision Tree, Naïve Bayes and Support Vector Machine. We also measured the performance, accuracy and precision of each hybrid method involved in the experiments by using different machine learning techniques. Text data was gathered from online news platforms for conducting the experiments.

This paper is structured as follows. Section II presents the literature review of related works on opinion mining techniques, with specific focus on information in the form of text in cyberspace. Section III describes the proposed theoretical framework developed based on an extensive literature review and background research of the technical opinion mining hybrid technique and national security, with particular focus on political security. Section IV highlights the evaluation process which includes the research design to extensively describe the research methodology and experimental analysis. A discussion on the experimental results is also provided in this section. Section V presents the results and analysis. The final section outlines the research summary, discusses future works related to this field as well as the conclusions achieved in this study.

II. LITERATURE REVIEW

A. OPINION MINING/SENTIMENT ANALYSIS

Opinion mining, or sentiment analysis, is the process of extracting opinions expressed in textual data or documents related to a particular objective. Natural Language Processing (NLP) can be applied in opinion mining. The trends of new types of text in social networks (such as blogs, online news, e-forums and e-commerce feedback) consist of text that contain people's opinions. The numerous subjects range from environment, politics and economics to technology, which can be shared across various online platforms. This massive number of subjective information can be used to compare opinions generated by people regarding products, customer feedback, brands, popularity monitoring, social media analysis, trending news and general moods [4]. Opinion mining can be classified into two sequential tasks, opinion detection and polarity detection, as well as the intensity of the opinions found in text fragments [5]. The key feature of opinion mining is that it evaluates a sample of text to comprehend the opinions developed within.

Research in opinion mining have mainly focused on the mechanism of labelling text to summarise the contents of emotion in texts and documents. Opinions are either labelled as positive or negative or may be classified as neutral when there is a lack of opinion in the text or when the viewpoint rests between two polarities. Several features can be used to identify opinion and polarity [6]. Assigning polarity to a text is an important process. Either positive, negative or neutral opinions are assigned to the text for opinion mining. Sentiment classification for opinion mining is an essential process to mine people's opinions based on text data. The common approaches used for sentiment classification are machine learning techniques, lexicon-based approach and the hybrid approach which combines two or more techniques [7].

B. LEXICON-BASED APPROACH

The lexicon-based approach uses a dictionary that includes word polarity [8]. The dictionary, also known as sentiment lexicon, is a compilation of sentiment terms. When a word appears in a text, it will be compared to a word in the dictionary to apply a sentiment score. The creation of a dictionary depends on the researchers since it can be done either automatically or manually with the help of seed words. Examples of existing sentiment lexicons include WordNet, SentiWordNet and NRC emotion lexicon [9]. The lexicon-based approach can be further classified into two types: dictionary-based approaches (based on dictionary words such as those found in WordNet or other sources) and corpus-based approaches (based on finding other opinion words in large corpus data) [10]. Such methods have been adopted for analysing news articles to determine people's sentiments [11].

When it comes to computation, the lexicon-based approach is usually an uncomplicated, effective method since it does not depend on any training data. The outcome can also be obtained much faster than the ML methods [12]. The implementation of this method is easy since it can classify people's opinions in text (such as positive or negative) by depending on dictionary scores listed in existing dictionaries. However, classifying emotions specific to a particular field is a problem since opinion mining cannot efficiently work without an existing library to compute the score of opinions and emotions in the texts. The existing libraries frequently used are the NRC word-emotion association lexicon (also known as the NRC emotion lexicon or Emolex) and the NRC emotion intensity lexicon. These libraries already contain the factor of emotion (known as affect intensity lexicons). For the emotion lexicon, the NRC word-emotion association lexicon was made in the English language to classify text into eight groups of emotions, including fear and anger. It differs from the NRC emotion intensity lexicon which can only classify text as either positive or negative emotions and sentiments. The NRC word-emotion association lexicon cannot tell if a piece of text is positive or negative since its English words list is linked to only eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise and trust) [2]. Since no dictionary for political security is available, we applied the existing NRC emotion lexicon in this research and combined this approach with selected machine learning techniques.

C. MACHINE LEARNING

Machine learning (ML) is a system that automatically acquires and integrates information due to its ability to learn through analysis, observation, training and experience. The ML outcomes from the analysis exhibit self-improvement, effectiveness and efficiency. For effective opinion mining, the ML framework using various techniques can be deployed for testing, interpreting and evaluating the acquired knowledge. Another benefit of ML algorithms is that they can be structured into a taxonomy based on the algorithm's expected outcome or the type of input available [13].

The ML method in opinion mining is also called supervised learning. This method produces a function, known as a label, that maps inputs to desired outputs. Several supervised learning methods can be applied such as the Naïve Bayes (NB) classification, Support Vector Machines (SVM) and Decision Tree (DT) classifiers [13]. These algorithms conduct their classification technique according to the training data. They learn and train themselves with respect to the differentiating attributes of the text. The performance of the classifier is tested using a test dataset. The machine learning process for sentiment analysis/opinion mining begins with the collection of datasets containing labelled sentences. These datasets may be noisy, therefore, they should be rehandled using various NLP techniques. Next, features relevant to opinion mining must be extracted. Finally, the classifier is trained and tested on the new sample data. These algorithms are extensively explained in the methodology section.

Most studies have applied the NB classification method for opinion mining or sentiment analysis due to two main reasons: 1) the NB classifier is widely known as a probabilistic classifier, and 2) it has the ability to describe text implementation [13]. Another widely used supervised machine learning algorithm for the purpose of classification is the SVM method. It is a statistical classification approach based on the maximisation of the margin between instances and the hyperplane separation. The SVM method is considered to be the best text classification method. In [14], the authors proposed an approach to acquire pre-labelled data from Twitter to train SVM classifiers. Twitter hashtags were used to judge the polarity of the tweet. To analyse the accuracy of the proposed technique, a test study on the classifier was conducted, yielding an accuracy result of eighty five percent (85%). The study also implemented the DT classifier, which is another supervised machine learning method. DT can classify data into different classes by recursively separating the feature space into two parts. It then assigns different classes according to which region in the divided space a sentence is, based on its features. These machine learning methods can be employed to enable computers to understand, analyse and classify the opinions expressed within texts.

D. HYBRID APPROACH

The hybrid approach is another method in opinion mining that combines two techniques or more to obtain accurate results [15]. Researchers frequently combine the ML method with the lexicon-based approach to optimize the advantages and reduce the drawbacks of both techniques in opinion mining. Although the ML method performs better as a classifier of text sentiment, it is not domain-dependent. To address this drawback, the hybrid approach has been proposed since combining various methods will achieve better results compared to simply using a standard approach based on a single method. Several researchers have applied the hybrid method by combining ML with the lexicon-based approach to increase the accuracy of ML. Neshan & Akbari found that the hybrid technique of combining ML with the lexicon-based approach enhanced the classification performance, turning it into a domain-dependent method. The hybrid approach can also be tailored to be more domain-dependent so that the outcome is accurate according to the respected domain [16].

Throughout the literature, several opinion mining methods (such as ML, lexicon-based approach and hybrid approach) have been extensively analysed and discussed. We found that the hybrid approach (combining the lexicon-based method with three machine learning techniques) can be a possible mechanism for evaluating people's opinions within text. The outcomes can be utilised to determine any possible security threats under the elements of political security. This will further secure and strengthen a state's national security from unwanted political security threats.

To classify and predict opinions within text, we employed a hybrid technique that combines the lexicon-based approach and NRC emotion lexicon with three selected machine learning techniques: Decision Tree, Naïve Bayes and Support Vector Machine.

Decision Tree: The decision tree is a hierarchical tree model structure of supervised learning that contains decision nodes to represent attributes and edges for denoting attribute values. A decision tree is quite an efficient method for creating classifiers from data. The representation of DT in the tree form allows it to construct decision rules that help classify the input of new data. The decision tree differentiates the textual dataset that has many features and checks the property from the root and vertex of the tree. All terminal vertexes are assigned a class label; 'pos' or 'neg'. Checking on property is the presence or absence of at least one word. The tree is partitioned until the least number of records exists. The decision tree algorithm is recursively done until the leaf nodes contain a certain minimum number of records used for classification.

Naïve Bayes: This is a well-known classifier in sentiment analysis/opinion mining works since it is the easiest to implement in data mining. The Naïve Bayes classifier considers each feature as different from another and employs the probability calculations based on the concept of a Bayesian approach. In a formula, prior probability is combined with conditional probability. The mathematical equation of this classifier is as follows:

$$P(x|Y) = \frac{P(Y|x)P(x)}{P(Y)} \tag{1}$$

Equation (1) display P(x) is the prior probability of a label or the likelihood that a label is observed. P(x|Y) is the prior probability that the feature set is classified as a label.

P(Y) is the prior probability that a given feature set has occurred. It shows that the naive prediction process acted as if all features are independent of each other.

For text classification, the Naïve Bayes classifier works by computing the posterior probability of a class based on the distribution of the words (features) in the document. We used the Multinomial NB, which is a Naïve Bayes function in python that works well for describing the discrete frequency of word counts in features.

Support Vector Machine: SVM is another machine learning classifier technique that involves a process of determining each raw fact as a dot for fixed dimensions of features. It would then choose a hyper-plane, one that maximises the margin between the two classes. The concept of SVM is the separation of learning space by a hyper plane, also known as a linear surface. This method considers the problem of linearity by applying a mathematical transformation to the learning space using kernel functions. When the transformation is complete, the data can be linearly separated and the optimal hyper plane, which is the normal distance of any data points, is the largest that can be found. For this research, the SVM python function is used as a text classifier for threat prediction.

E. NATIONAL SECURITY

National security denotes the protection of the state and its people [17]. National security is a situation in which the country and culture are protected from threats, such as violent invasion, political repression and economic coercion [18]. Protecting the country and its people from threats and other possible dangers by maintaining armed powers, as well as protecting state secrets, have been the highest priority of national security itself. Balzacq, Bahadur, Thakuri & Kshetri outlined a list of national security elements, such as military security, political security, cybersecurity, human security, economy security, homeland security, energy and natural resource security, border security, food security and health security [19], [20], [21]. These elements are illustrated in Fig. 1.

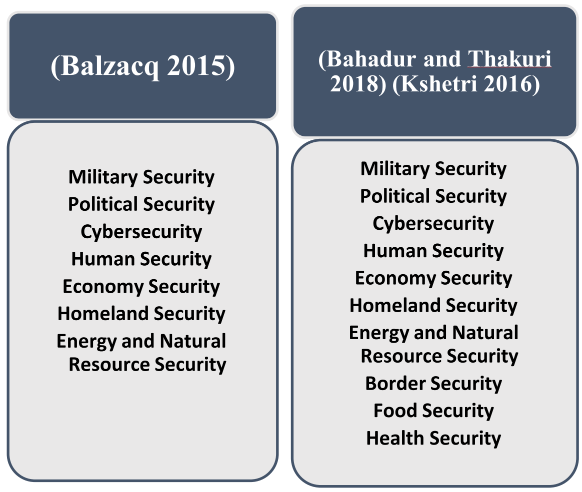

FIGURE 1. National security elements.

To maintain the stability of national security, any problems related to social, political and economic issues must be constantly monitored according to the eight elements of national security: military security, political security, human security, cybersecurity, economic security, homeland security, environmental security as well as energy and natural resources security [19]. Threats associated with a national security risk will have an impact on the nation's security and stability. If there are any direct or indirect threats to a nation, the government should mobilise their national security structure to ensure that such threats are managed appropriately. The nation will face threats that weaken the state of national security, for instance, imbalanced government, natural disasters, terrorism, cybercrime and illnesses. The growth of the internet technology is also one of the biggest threats to national security due to its negative impact [22].

The advances of the internet have made cyberspace a platform for spreading fake news and negative rumours which can threaten a country [23]. People can spread rumours, hate speech and fake news using any type of social network sites, such as social media, blogs and online news. The information online will be accepted as true since the negative posts written in cyberspace will spread without verifying the facts [24]. These scenarios have the potential to influence public opinion or sentiment and cause unwanted events, such as protests and social wars, which can jeopardise people's faith in the government and consequently threaten national security. The influence of human emotions within internet text can exert a negative impact on national security.

Hence, there is a need to evaluate people's opinions or sentiments regarding text in cyberspace to avoid excessively negative emotions that can cause chaos. Research have shown that individual needs, opinions, emotions and behaviours can be revealed from analysing textual data [25]. Thus, a mechanism to monitor public opinion and emotions related to military security, political security, human security, homeland security, cybersecurity, economic security, environmental security as well as energy and natural resources security is crucial to maintain the stability of the country.

F. NATIONAL SECURITY THREAT CLASSIFICATION FOR POLITICAL SECURITY

A threat is an event that must be addressed, solved or avoided due to the dangers it may pose, especially when related to national security threats. All national security elements have their criteria and threats. To align with the present trending world, the element of national security that acts as a significant threat to the state is political security.

Political stability is of vital interest to national security since several threats from the political aspect can harm the country and weaken its political security, such as violations of the rule of law caused by tensions between communities. A stable political security can strengthen the state's national security. A study had noted that national security must be viewed in terms of political security to protect and promote national security goals and objectives [26].

Threats that shake the stability of a country's political security would be political violence, political upheaval and political repression since they jeopardise the country's peace. To date, political issues can be easily expanded to cyberspace due to the sophisticated use of the internet, posing a threat to the political atmosphere of the country [27]. Threats must be managed to avoid deteriorating levels of political security.

This study addresses emotion as the related attribute to political security threats since emotion can influence the sentiments or opinions that affect the political dimension. Figure 2 portrays the relationship between emotion and political security threats based on the study by Coan, Merolla & Zechmeister who stated that 'fear' and 'anger' can influence changing political behaviours and public opinion [28].

One of the dominant emotions is anger. When citizens feel dissatisfied with political institutions, they show their anger by protesting or causing political upheavals, such as coups or riots. Passarelli & Tabellini found that riots are entirely spontaneous and exclusively motivated by emotions. Rioting is also referred to as an action to "take revenge" or display anger for an unfair outcome [29], [30].

Political upheaval can occur when countries experience changes in social attitudes and perspectives, causing unrest. Changes in politics and political movements can both disrupt economies and weaken national political security. Political violence is also a political security threat. However, political terrorism and civil war are different forms of political violence since they can elicit fear from citizens regarding a situation. A study by Sousa revealed that political violence can damage community culture and cause mass killings and displacement, destruction of meaningful places as well as control of space and movement.

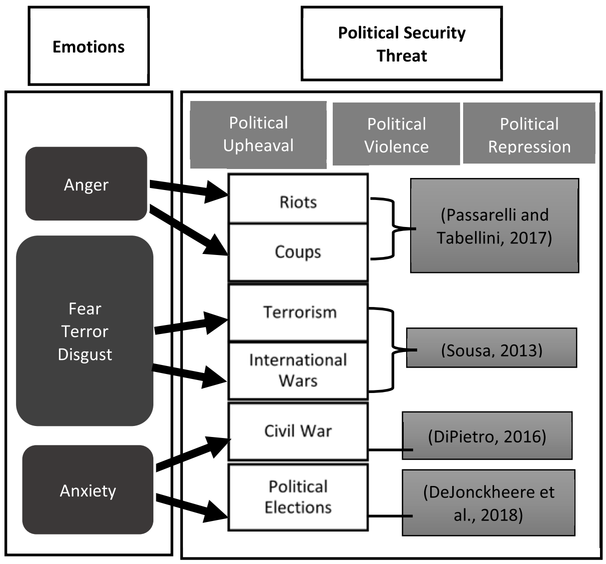

FIGURE 2. The relationship between emotion, sentiment and political security threats.

Consequently, people will either feel fear, disgust or terror when it comes to such situations [31]. According to Jonathan, the expression of anger during political communication can be levelled at a specific opponent or at a particular scenario or event that had caused public dissatisfaction [32]. For instance, during an election, anxiety is the emotion eventually released by the community when they feel worried about any injustice or misunderstanding that may have occurred during the election process.

A recent study by De Jonckheere et al. stated that feelings of anxiety influence people during election season, especially among young women [33]. Political repression can occur when feelings of anxiety become too extreme, thus causing the public to distrust their leaders and the system itself. It can also lead to fights between society members which will simultaneously disrupt a country's political stability or lead to a civil war. Civil conflict between neighbouring countries can also lead to repression within a country, even if they are not experiencing civil war [34]. When fake and negative sentiments/opinions circulate online, people can easily be influenced since they generally believe what they see online. Problems will arise when a community also believes in the

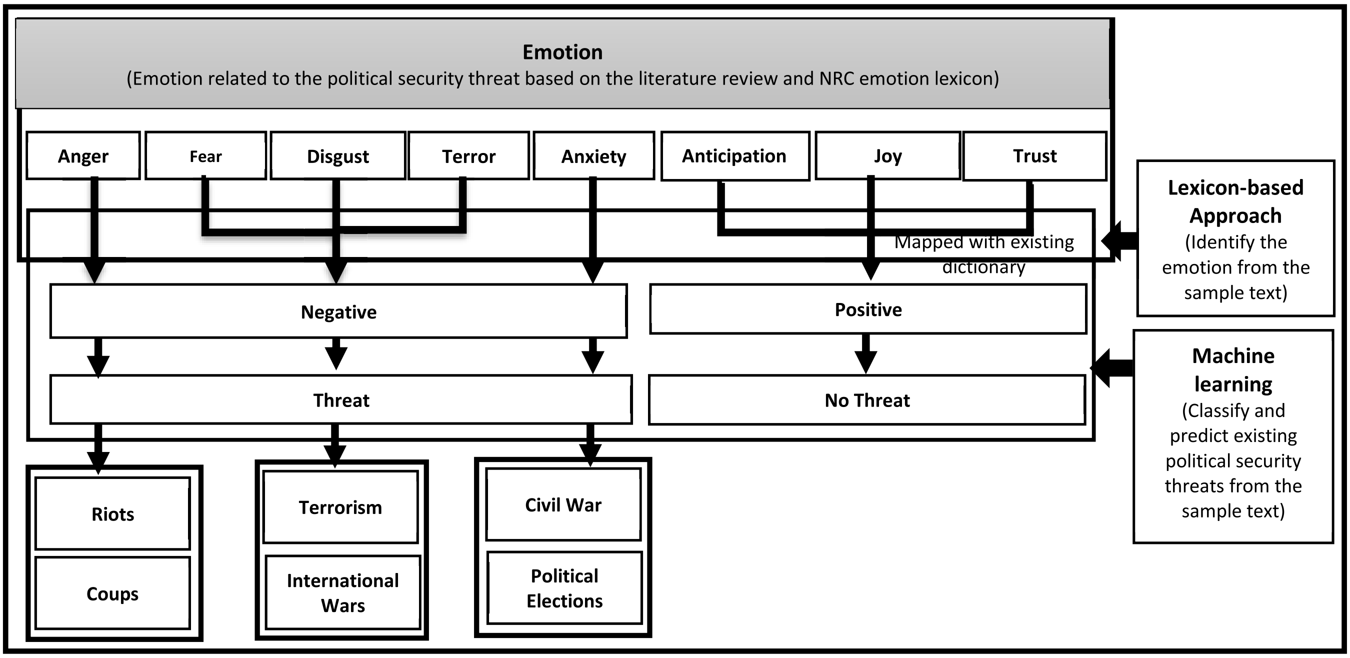

FIGURE 3. Theoretical framework of political security threat prediction using the hybrid lexicon-based approach and machine learning.

circulating negative sentiments, which will create confusion as such emotions spread [35].

III. THEORETICAL FRAMEWORK IN THE PREDICTION OF POLITICAL SECURITY THREAT

Based on the discussion in the previous section, opinion or sentiment monitoring is crucially needed to control risks that may occur due to threat-related political security events. Emotion is a variable that influences opinion. Thus, this research proposes a new theoretical framework for political security threat prediction based on opinion embedded with emotion. We suggest that political security threat-related events are influenced by emotions stemming from sentiments/opinions. To validate our proposed framework, we established an experimental design using the hybrid lexicon-based approach and selected machine learning techniques. Figure. 3 displays the proposed theoretical framework of this research.

We propose that emotion is a key variable in determining opinions or sentiments. The literature review has shown that the listed emotions (anger, fear, disgust, terror and anxiety) can act as emotion indicators in determining the existence of political security threats in textual data. We propose that the listed emotions, as determined by the literature, is related to the political security domain and can trigger political events such as riots, coups, terrorism, international war, civil war and political elections, which can lead to negative opinions/sentiments. Opinions or sentiments can be mined to conduct threat predictions in the political security domain. The mining process can apply various techniques, as suggested throughout the literature.

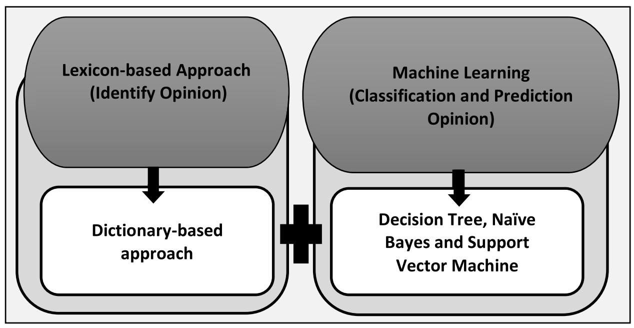

FIGURE 4. Opinion mining concept using the hybrid technique for political security.

A. RESEARCH DESIGN

To validate our proposed theoretical framework, the experiments were designed based on the concepts shown in Figure. 4. For the experiments, we employed the hybrid technique (a combination of machine learning techniques and the lexicon-based approach) to predict political security threats.

The technique illustrated in Fig. 4 is the combination of the lexicon-based approach and selected machine learning techniques. This will demonstrate the relationship between people's emotions, opinions/sentiments and political security threats by classifying and predicting the polarity of both emotion and opinion related to threats.

Our hybrid method uses a dictionary for the lexicon-based approach and machine learning techniques which include Decision Tree, Naïve Bayes and Support Vector Machine. They were included for the classification process and opinion prediction of the input sample text. The results were evaluated by conducting precision, recall and accuracy tests. These experiments will determine and extract the opinions of the textual dataset. Opinions act as the factor in predicting the existence of threats based on political security.

B. RESEARCH METHODOLOGY

The developed research methodology includes the concepts, algorithms, data collection, experiment tools and analysis results. Figure. 5 presents the research methodology of this study. The experimental design includes analysing results according to the selected national security domain, which is political security. Data was then collected from online news and labelled using self-developed tools. The lexicon-based approach was implemented to label the opinion of the input text. A machine learning algorithm was then applied to the textual dataset with the opinions already labelled for each text for classification. A prediction of the new sample text is then accomplished by calculating whether or not sentences are embedded with either positive or negative sentiments or opinions.

To determine the accuracy of each hybrid method used, evaluation was conducted by employing the metric performance evaluation to determine accuracy, precision and recall.

C. EXPERIMENTAL ANALYSIS

In this phase, the mechanism and tools used for the experimental analysis process are accomplished. The Python programming language was used in the experiments for programming according to the hybrid technique applied in this research. This is to generate, store, and classify the opinions/sentiments in the dataset. For the lexicon-based approach, the NRC emotion dictionary lexicon was used. For machine learning, the decision tree, Naïve Bayes and support vector machine were employed for opinion prediction. The experiment was executed throughout five phases: i) data collection, ii) data pre-processing, iii) opinion extraction, iv) opinion prediction and v) evaluation and validation. Figure. 6 displays the five phases of the experimental design.

1) DATA COLLECTION

In the data collection phase, the test dataset for the experiment was manually collected from online news. Three online news platforms were selected: The Star, New Straits Times (NST) and Free Malaysia Today (FMT). These sample texts serve as the dataset to evaluate the proposed hybrid approach in this research. The sample texts were taken in the form of sentences from the news articles. Taking the sample text in sentence form will prevent complicated or false extraction of opinions.

Extracting sentences would be clearer and more specific than extracting entire paragraphs. The collected sample sentences are raw since the opinions within were not labelled. Next, the collected sentences were inputted in an excel document.

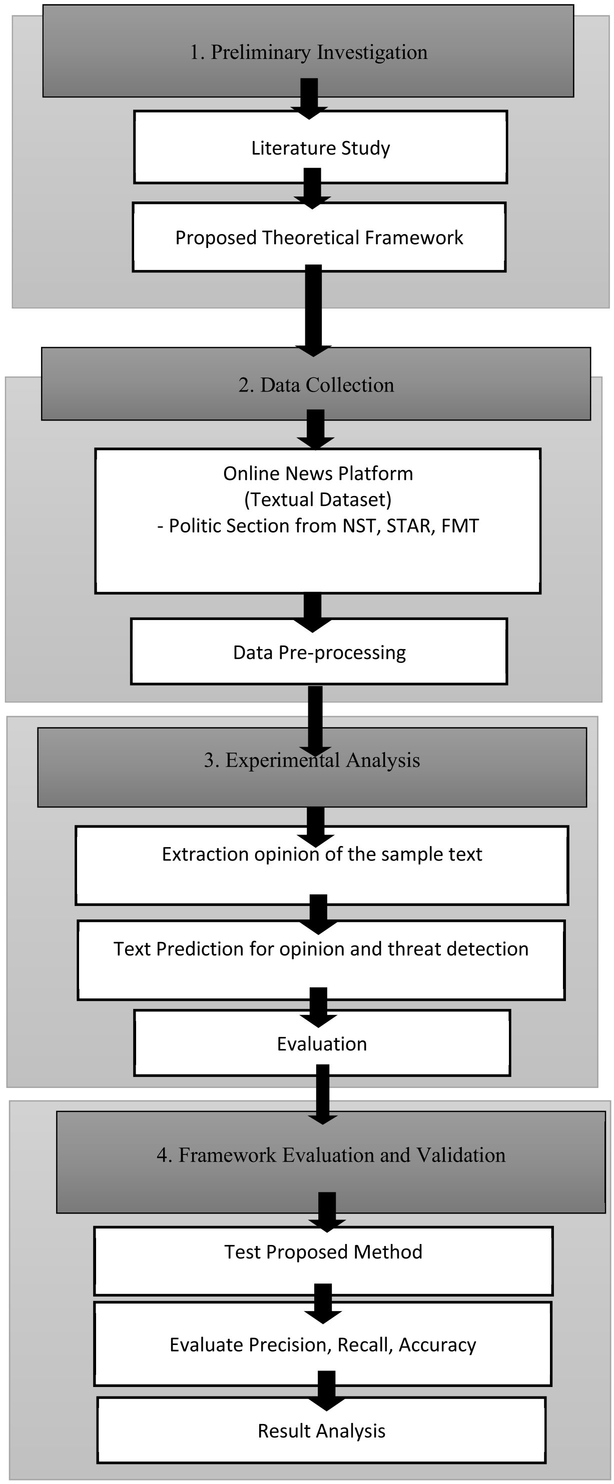

FIGURE 5. Research methodology.

2) DATA PRE-PROCESSING

After the data collection process, the data pre-processing phase was carried out to extract the relevant text in preparation for the textual dataset. Pre-processing data is necessary to eliminate noise, inconsistent or incomplete data. The collected sample text from the online platform consists of irrelevant symbols, such as hashtags and special characters, that

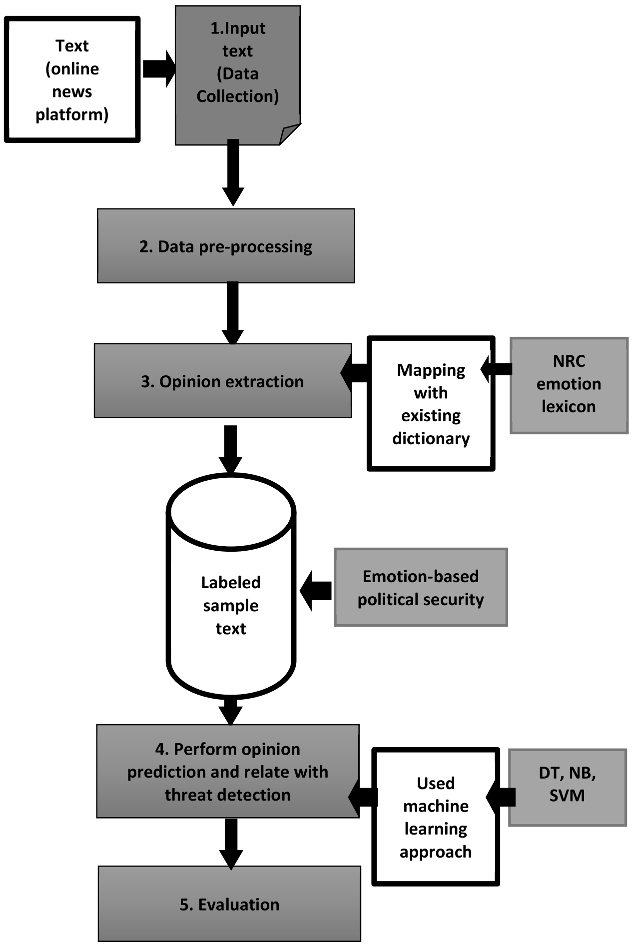

FIGURE 6. Experimental design.

TABLE 1. Sample sentences before pre-processing.

Sentences
The minister should understand that our lives now revolve around the TV and computer screens, and chairs are as important as our eyes.
"Politics must be set aside as the war on Covid-19 is at a critical stage where it doesn't only affect lives but also this country's future," he said.
First, one may ask, what exactly is money politics? As there is no set definition, each individual has a different answer.
"Every person deserves their politicians, who are often just a reflection of society's ugly true self. As some wise person says, 'you are never really caught in the jam. You are the jam!'" he said.
"The people have lost confidence that the council can solve the issue amicably," he said.

must be removed since they are considered to be noise. Data pre-processing includes removing the URL, removing special characters, removing hash symbols as well as removing stop words and word stems. The sentences before and after pre-processing are shown in Tables 1 and 2. The opinions are extracted from these sentences in the next phase.

TABLE 2. Sample sentences after pre-processing.

Sentences
minister understand lives revolve around tv computer screens chairs important eyes
politics must set aside war covid19 critical stage doesnt affect lives also countrys future said
first one may ask exactly money politics set definition individual different answer
every person deserves politicians often reflection society ugly true self wise person says never really caught jam jam said
people lost confidence council solve issue amicably said

3) OPINION/SENTIMENT EXTRACTION

In this phase, the lexicon-based approach is implemented for opinion extraction using the dictionary-based technique to extract the emotions within the dataset. The NRC emotion lexicon was employed for extracting the opinion within the context.

NRC: NRC is an emotion lexicon resource that stands for National Research Council of Canada. This lexicon was manually created by Mohammad and Turney to conduct a tag process using the Mechanical Turk and WordNetAffect. A collection of more than 14,000 English words was distributed to the people to annotate the emotion of the word based on Plutchik's wheel of emotion in the Mechanical Turk platform. Next, the words were analysed based on the emotion categories: fear, sadness, anger, disgust, trust, anticipation, surprise and joy. The emotions were also categorised as either positive or negative words. Each word was associated with binary scores (0 or 1).

The emotion values of the text were calculated by linking them with the score of specific words in that lexicon, 0 being the minimum value of emotion while 1 is the maximum value of emotion. Each word was marked for emotion and sentiment classification. The emotional score was then labelled for each sentence. Table 3 presents the algorithm of this process, and Table 4 displays the output of sentences labelled with an emotion as well as the emotional score at this phase.

The sentences were classified as either positive or negative based on the emotion extracted from them. Table 5 provides the sentences labelled with either positive or negative opinions. Several emotions were listed in the positive and negative class.

Table 6 below presents the results of the lexicon-based approach.

Table 6 shows the resulting output when applying the NRC emotion lexicon as the opinion extraction. A file with a total of 250 online news texts is provided. We achieved a total of 163 positive texts and 87 negative texts. The extracted positive and negative opinions are the indicators in the classification of determining whether or not the sample texts contain any threats. Fig. 7 below lists emotions as positive and negative classes to classify whether the sample sentences pose any threats to political security.

TABLE 3. Algorithm of opinion extraction.

```
Algorithm 1: Performing opinion mining (text opinion extraction using the lexicon-based approach)
Input: sample text from online news
Output: Text label with polarity score and emotion
Start:
Input Data: Sample text, import the sentiment lexicon (NRC emotion lexicon)
Result: "Text Opinion Extraction" and polarity score
1 Opinion Extraction () ← File
2.For each sample text line then do
        Initialize the emotion of sample text in sentiment Lexicon
        if sample text = No emotion value label to it then
        Sync Emotion word = emotion dataset
        Positive Emotion  ←Happy (example)
        Update the text label emotion
        Else
            if sample text = Negative emotion then
                Sync Emotion word = emotion dataset
                Emotion  ←Sad (example)
                Update the text label emotion
                  else
          End
        End
    End
End
```

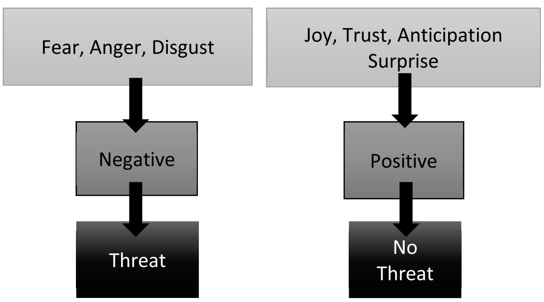

FIGURE 7. Opinion-labelled classes.

The results obtained from the NRC emotion lexicon phase can extract the positive and negative opinions from the text. The sample text in the textual dataset was updated with the labelled opinions. This latest dataset was used in the next phase for acquiring accurate results in opinion prediction. The extracted results were stored in different files and treated as inputs for the next machine learning approach.

4) OPINION/SENTIMENT PREDICTION FOR POLITICAL SECURITY USING HYBRID TECHNIQUES

In this phase, opinion mining was accomplished to predict and classify opinions within the text as either positive or negative. This stage was conducted for the purpose of validating the proposed approach by predicting whether or not the opinion of the new input sample text possesses any threats to political security. The decision tree was used in this phase to calculate and predict the opinion for each sample text. The output from this phase is labelled as either positive or negative sentiment. From here, we can identify whether sentences possess any threats based on the definition of political security threat, as previously discussed. The steps applied in this phase are as follows:

TABLE 4. Example of labelled emotions of sample sentences.

Sentence	Emotion Score	Emotion
The minister should understand that our lives now revolve around the TV and computer screens, and chairs are as important as our eyes	0.5	Trust
"Politics must be set aside as the war on Covid-19 is at a critical stage where it doesn't only affect lives but also this country's future, " he said.	0.333333333333333	Fear
That being said, a political financing legislation, while much needed, is apparently unpopular among politicians and political parties as it would severely restrict their ability to source for funds to ensure their and their party's survival	0.333333333333333	Sadness
Anas explained that Rasuah (corruption) Busters was not trying to reshape Malaysian politics but Malaysian culture. He said they were trying to create a new platform, new hope and new belief that we can reverse the rasuah (corruption) culture	0.333333333333333	Joy
Slaughtering of chickens in markets is prohibited. The process along with distribution are supposed to be done through a centralised distribution house	0.2	Anger
There's no better description of nation today than to say we are the "sick man of Asia". Sick, because of the Covid-19 pandemic, sicker because our politicians make us sick.	0.333333333333333	Disgust

- *Feature Extraction:*

The feature acts as a computable characteristic of an attribute. The textual dataset classified each sentence as positive or negative using the NRC emotion lexicon. Each sentence was considered as a simple document. Positive and negative texts were labelled as 'pos' and 'neg' for each document. The documents were divided into training and testing data. The TF-IDF (Term Frequency-Inverse Document Frequency) was then used as the feature extraction method to weigh the training and testing data. The TF-IDF is one of many numerical statistics which has the function to reflect the importance of a term in a document inside a corpus or collection. In this step, the TF-IDF only assigns weight scores on the occurrence frequency of sentence data in the document and not the labels. After the feature extraction of data, the next step is the opinion classification and prediction. This stage trains and implements machine learning algorithms.

TABLE 5. Example of labelled sample sentences according to their opinion/sentiment.

Sentence	Emotion	Sentiment/ Opinion class
The minister should understand that our lives now revolve around the TV and computer screens, and chairs are as important as our eyes	Trust	Positive
"Politics must be set aside as the war on Covid-19 is at a critical stage where it doesn't only affect lives but also this country's future, " he said.	Fear	Negative
That being said, a political financing legislation, while much needed, is apparently unpopular among politicians and political parties as it would severely restrict their ability to source for funds to ensure their and their party's survival	Sadness	Negative
Anas explained that Rasuah Busters was not trying to reshape Malaysian politics but Malaysian culture. He said they were trying to create a new platform, new hope and new belief that we can reverse the rasuah (corruption) culture	Joy	Positive
Slaughtering of chickens in markets is prohibited. The process along with distribution are supposed to be done through a centralised distribution house	Anger	Negative
There's no better description of nation today than to say we are the "sick man of Asia". Sick, because of the Covid-19 pandemic, sicker because our politicians make us sick.	Disgust	Positive

TABLE 6. Result of lexicon-based approach.

Method	Total Text	Positive	Negative
Lexicon based Approach	250	163	87

- *Classify and predict:*

This stage processes the classification and prediction of text opinions and existing political security threats of the new sample texts. The classification machine learning algorithms selected for classifying and predicting the opinions within text are decision tree (DT), Naïve Bayes (NB) and support vector machine (SVM). Fig. 8 displays the flowchart of the machine learning techniques employed in this research.

For this research, machine learning algorithm is recursively conducted until the prediction result is obtained for the class, as shown in Table 7 This method is used in classifying and predicting the opinions of text. After completing the classification and prediction processes, one more step is needed to determine their quality, namely, the evaluation of results. At this stage, the performance of conducted calculations are tested with the performance metrics.

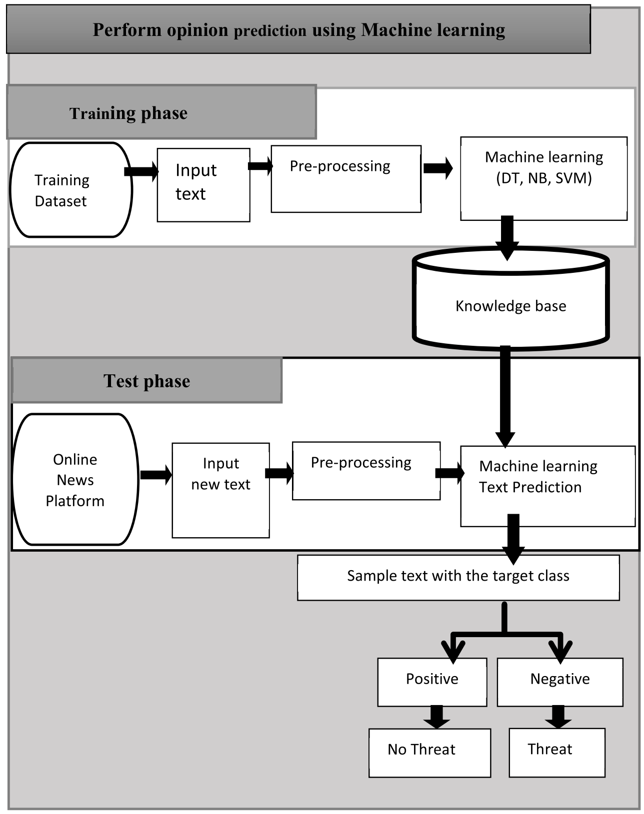

FIGURE 8. Flowchart of the machine learning techniques.

TABLE 7. Algorithm of opinion prediction.

Algorithm 2: Perform opinion prediction of proposed approach using the machine learning algorithm

Output: Text label with polarity score and opinion
Start:
Input Data: Test set (new sample text collected in excel) and Document of the training set (Dataset of result from algorithm 1)
1 Predict Opinion classification ()⟵ File
Read the training dataset
2.For each sample text line in test set do
Text (positive emotion) calculate (positive emotion)
Text (negative emotion) calculate (negative emotion)
If Text (positive emotion) > Text (negative emotion) then
Text is positive emotion
end if
If Text (positive emotion) < Text (negative emotion) then
Text is a positive emotion
end if
end for
end

TABLE 8. Example of predicted results of the hybrid approach.

Sentence	Emotion	Opinion/Sentiment			Prediction Class		
		DT	NB	SVM	DT	NB	SVM
The minister should understand that our lives now revolve around the TV and computer screens, and chairs are as important as our eyes	Trust	Positive	Positive	Positive	No Threat	No Threat	No Threat
"Politics must be set aside as the war on Covid-19 is at a critical stage where it doesn't only affect lives but also this country's future," he said.	Fear	Negative	Positive	Positive	Threat	No Threat	No Threat
My curiosity was piqued. Politics in the time of Covid-19 has turned me off. There is to much politicking when the country need to be focused on fighting this pandemic	Anger	Negative	Positive	Positive	Threat	No Threat	No Threat
Anas explained that Rasuah Busters was not trying to reshape Malaysian politics but Malaysian culture. He said they were trying to create a new platform, new hope and new belief that we can reverse the rasuah (corruption) culture	Joy	Positive	Positive	Negative	No Threat	No Threat	No Threat

5) EVALUATION TO DETERMINE THE PERFORMANCE OF THE HYBRID TECHNIQUE

The hybrid technique used in the experiment is evaluated in this final stage. The evaluation process involves checking the accuracy, precision and recall of the results from the proposed hybrid approach, which is a combination of the lexicon-based approach and several machine learning techniques used in the experiments: decision tree, Naïve Bayes and support vector machine.

The evaluation process began after labelling data with a class of either positive or negative tags and removing the imbalanced proportion of classes. A random sample of 250 sentences was taken to train and test the dataset with the proposed method results, which includes DT, NB and SVM classifiers. Next, a confusion matrix was implemented to calculate the accuracy value of ML classifiers such as precision, recall and accuracy. This is to determine which algorithm produces better accuracy values based on the labels in the training data. A certain formula was used to calculate the accuracy, precision and recall values for the performance evaluation of these classifiers.

Equation (2) is to calculate the accuracy, we applied True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

The precision value was calculated using the TP and FP formula, as shown in "(3)", below.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

The formula to calculate the recall values using TP and FN is refer to "(4)", follows:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

A confusion matrix is crucial to compare algorithms and determine which algorithm is better in specific cases. By analysing the classified scores, this research can conclude the best hybrid method to employ for our proposed framework. The accuracy was also calculated.

IV. RESULT

This research proposed a theoretical framework for political security threat prediction using the hybrid of lexicon-based approach and machine learning techniques. Table 8 provides an example of the predicted results using the hybrid approach.

Table 8 presents an example of text prediction using the hybrid approach. The lexicon-based approach with the decision tree classifier presented more accurate results compared to the other combinations (NB and SVM). It showed the right predicted label of opinion/sentiment and the existence of threats according to the political security domain.

The results of the hybrid approach experiment were evaluated by comparing the result values of accuracy, precision and recall with each combination of the lexicon-based approach and the three machine learning techniques (DT, NB and SVM). The values of accuracy, precision and recall are important for measuring and comparing the performance of one method with another. Positive and negative files are used as training samples for new test data, which have formed the base for feature extraction using machine learning techniques. The evaluation results of our test data are shown in the following table. Table 9 displays the performance measure of the hybrid lexicon-based approach and decision tree classifier.

Table 9 reveals that the performance measure of the lexicon-based approach combined with a decision tree classifier achieved 69% overall accuracy and correctly classifying 69% of instances. The model had a precision score of 88%, precisely predicting a particular class 88% of the time, and a recall score of 67%, accurately identifying 67% of instances belonging to a particular class. The increase in the number of false negatives suggests that the model is missing some instances that belong to a particular class, resulting in a decrease in recall. However, the decrease in false positives indicates that the model is making fewer incorrect

TABLE 9. Lexicon-based approach + decision tree performance.

	False Negative	False Positive
True Negative	21.33%	6.67%
True Positive	24.00%	48.00%
Precision	88%	
Recall	67 %	
Accuracy	69%	

predictions, resulting in an increase in precision. When precision is higher, it means that the model is more confident in its predictions, but it also suggests that some instances that belong to a particular class were not identified, resulting in a lower recall score. In other words, the model is more likely to predict the correct class, but it may miss some instances that belong to that class.

The performance of the lexicon-based approach with the Naïve Bayes is displayed in Table 10 below.

TABLE 10. Lexicon-based approach + NAÏVE bayes performance.

	False Negative	False Positive
True Negative	2.67%	42.76%
True Positive	0.00%	54.67%
Precision	56%	
Recall	100%	
Accuracy	57%	

Table 10 reveals that the combination of the lexicon-based approach with the NB classifier achieved an overall accuracy of 57.3%, which is quite low compared to the achievement of DT. The model has high recall and low precision, indicating that it correctly identifies most positive instances (true threats) but also incorrectly identifies many negative instances (false positive). This means that the model has a tendency to classify non-threatening text as threatening, leading to a decrease in precision. This study also notes that the model needs to be improved to reduce the false positive rate and increase precision. The NB classifier suffered the most in precision value while classifying the text into a ‘‘Negative’’ class. The performance of the lexicon-based approach with SVM is presented in Table 11 below.

TABLE 11. Lexicon-based approach + support vector machine performance.

	False Negative	False Positive
True Negative	6.67%	0.00%
True Positive	38.76%	54.67
Precision	59%	
Recall	100.0 %	
Accuracy	61%	

Table 11 shows that the combination of the lexicon-based approach and the SVM classifier achieved an overall accuracy of 61%, which is lower than DT but higher than NB. The evaluation result showed a higher recall value than the precision value, with the recall value being 100%. A higher recall value shows that the model is properly identifying all the positive instances in the data, but it may also be including a lot of false positives in its predictions. In other words, the model is capturing a larger proportion of positive instances but is also making more errors by incorrectly classifying negative instances as positive. Overall, the trade-off between precision and recall should be considered based on the specific needs of the task at hand. If identifying all instances of a particular class is important, then a higher recall score may be more desirable, even if it results in lower precision. On the other hand, if minimizing false positives is crucial, then a higher precision score may be more important, even if it results in a lower recall. On average, it can be concluded that the lexicon-based approach with the decision tree classifier is the best technique for opinion prediction in the political security domain, based on its high average accuracy of 69%, outperforming other machine learning algorithms, with high precision and low recall values. This approach is also effective in predicting threat classes correctly and providing stable and accurate predictions of the positivity or negativity of a sentence based on opinion words.

V. CONCLUSION

The research introduced a theoretical framework for predicting political security threats using a hybrid approach of lexicon-based analysis and machine learning techniques. This framework is designed to analyze people’s opinions on the national security domain, with a specific focus on the political security element. The research aims to enhance opinion mining in the national security domain, and it includes opinion mining and national security elements specific to political security to create a multi-research domain study.

The research successfully demonstrated the relationship between emotions, opinions, sentiment, and political security threats in cyberspace. The research presents a new theoretical framework that utilizes the lexicon-based approach and machine learning for the emotional assessment of text in the national security domain, specifically for the political security element. The study concludes that the combination of the lexicon-based approach with the decision tree classifier is the best hybrid approach method for detecting political security threats based on emotions embedded within online news text.

As future work, a performance analysis of the proposed method using a massive dataset for this method will be conducted. This future work aims to create new research fields to incorporate opinion mining in the domain of national security, particularly for political security elements. This future work is expected to contribute to the establishment of a more efficient and effective hybrid methodology that can be used in practice to predict political security threats based on emotions embedded within textual data. Overall, this research has the potential to significantly improve the ability to predict and prevent political security threats, which is critical for ensuring national security.

ACKNOWLEDGMENT

The authors fully acknowledged UPNM and the Ministry of Higher Education Malaysia (MOHE) for the approved fund, which makes this research viable and effective.

REFERENCES

[1] J. R. Clapper, "Statement for the record: Worldwide threat assessment of the us intelligence community," Office Director Nat. Intell., Congressional Testimonies 2015, USA, 2015. [Online]. Available: https://www.dni.gov/files/SFR-DirNCTCSHSGACHearing8Oct.pdf
[2] N. A. M. Razali et al., "Opinion mining for national security: Techniques, domain applications, challenges and research opportunities," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00536-5.
[3] S. Dorle, "Sentiment analysis methods and approach: Survey," *Int. J. Innov. Comput. Sci. Eng.*, vol. 4, no. 6, pp. 1–5, Dec. 2017, [Online]. Available: http://www.ijicse.in/index.php/ijicse/article/view/134
[4] A. Balahur, R. Steinberger, E. Van Der Goot, B. Pouliquen, and M. Kabadjov, "Opinion mining on newspaper quotations," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol.*, Sep. 2009, pp. 523–526, doi: 10.1109/WI-IAT.2009.340.
[5] B. Seerat, "Opinion mining: Issues and challenges(A survey)," *Int. J. Comput. Appl.*, vol. 49, no. 9, pp. 42–51, 2012, doi: 10.5120/7658-0762.
[6] P. Barnaghi, J. G. Breslin, I. D. A. B. Park, and L. Dangan, "Opinion mining and sentiment polarity on Twitter and correlation between events and sentiment," in *Proc. IEEE 2nd Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar./Apr. 2016, pp. 52–57, doi: 10.1109.
[7] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.
[8] G. Isabelle, W. Maharani, and I. Asror, "Analysis on opinion mining using combining lexicon-based method and multinomial Naïve Bayes," in *Proc. Int. Conf. Ind. Enterprise Syst. Eng.*, vol. 2, 2019, pp. 214–219, doi: 10.2991/icoiese-18.2019.38.
[9] H. Zhang, W. Gan, and B. Jiang, "Machine learning and lexicon based methods for sentiment classification: A survey," in *Proc. 11th Web Inf. Syst. Appl. Conf.*, Sep. 2014, pp. 262–265, doi: 10.1109/WISA.2014.55.
[10] M. P. Ashna and A. K. Sunny, "A study on sentiment analysis in Malayalam language," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 6, no. Special Issue 3, pp. 88–93, 2017, doi: 10.17148/IJARCCE.
[11] S. Taj, B. B. Shaikh, and A. Fatemah Meghji, "Sentiment analysis of news articles: A lexicon based approach," in *Proc. 2nd Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Jan. 2019, doi: 10.1109/ICOMET.2019.8673428.
[12] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proc. Int. Conf. Web Search Web Data Mining (WSDM)*, 2008, pp. 231–239, doi: 10.1145/1341531.1341561.
[13] B. Saberi and S. Saad, "Sentiment analysis or opinion mining: A review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1660–1666, 2017, doi: 10.18517/ijaseit.7.5.2137.
[14] W. A. Zgheib and A. M. Barbar, "A study using support vector machines to classify the sentiments of tweets," *Int. J. Comput. Appl.*, vol. 170, no. 2, pp. 8–12, Jul. 2017, doi: 10.5120/ijca2017914690.
[15] G. Vaitheeswaran and D. L. Arockiam, "Combining lexicon and machine learning method to enhance the accuracy of sentiment analysis on big data," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 1, pp. 306–311, 2016, doi: 211541386.
[16] A. D'Andrea, F. Ferri, P. Grifoni, and T. Guzzo, "Approaches, tools and applications for sentiment analysis implementation," *Int. J. Comput. Appl.*, vol. 125, no. 3, pp. 26–33, Sep. 2015, doi: 10.5120/ijca2015905866.
[17] S. L. Cosby, "Human security concept: The root of U.S. national security and foreign policy," United States Mar. Corps, USA, 2009.
[18] S. Stoyko and G. S. Rakovski, "Management, security and the system of national security," Military Acad., Sofia, Bulgaria, 2011.
[19] T. Balzacq, "What is national security?" *Revue Internationale et Stratégique*, vol. 52, no. 4, pp. 33–50, 2015, doi: 10.3917/ris.052.0033.
[20] B. Bahadur and S. Thakuri, "Human security?: Concept, dimensions & challenges," Inst. Crisis Manage. Stud., Kathmandu, Nepal, 2018.
[21] N. Kshetri, *The Quest to Cyber Superiority: Cybersecurity Regulations, Frameworks, and Strategies of Major Economies*, 1st ed. Cham, Switzerland: Springer, 2016.
[22] P. Cornish, R. Hughes, and D. Livingstone, *Cyberspace and the National Security of the United Kingdom*. London, U.K.: Chatham House, Mar. 2009, p. 46.
[23] S. Cordey, "Cyber influence operations: An overview and comparative analysis," Tech. Rep., Oct. 2019.
[24] M. Yassine and H. Hajj, "A framework for emotion mining from text in online social networks," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Dec. 2010, pp. 1136–1142, doi: 10.1109/ICDMW.2010.75.
[25] A. F. Ab Nasir et al., "Text-based emotion prediction system using machine learning approach," in *Proc. IOP Conf. Ser. Mater. Sci. Eng.* IOP Publishing, vol. 769, no. 1. 2020, doi: 10.1088/1757-899X/769/1/012022.
[26] D. Enclave, R. Tula, and R. Marg, *Institute for Defence Studies and Analyses (IDSA)*, no. 1. 2019.
[27] R. Sandoval-Almazan and J. R. Gil-Garcia, "Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements," *Government Inf. Quart.*, vol. 31, no. 3, pp. 365–378, 2014.
[28] T. G. Coan, J. L. Merolla, and E. J. Zechmeister, "Emotional responses to human security threats: Evidence from a national experiment," Tech. Rep., 2012, pp. 1–26.
[29] F. Passarelli and G. Tabellini, "Emotions and political unrest," *J. Political Economy*, vol. 125, no. 3, pp. 903–946, Jun. 2017, doi: 10.1086/691700.
[30] P. Battigalli, M. Dufwenberg, and A. Smith. (2015). *Frustration and Anger in Games*. [Online]. Available: https://econpapers.repec.org/RePEc:igi:igierp:539
[31] C. A. Sousa, "Political violence, collective functioning and health: A review of the literature," *Med., Conflict Survival*, vol. 29, no. 3, pp. 169–197, Sep. 2013, doi: 10.1080/13623699.2013.813109.
[32] J. Van Riet, G. Schaap, M. Kleemans, and S. Lecheler, "On different sides: Investigating the persuasive effects of anger expression in political news messages," *Political Psychol.*, vol. 40, no. 4, pp. 837–857, 2019, doi: 10.1111/pops.12554.
[33] M. DeJonckheere, A. Fisher, and T. Chang, "How has the presidential election affected young Americans?" *Child Adolescent Psychiatry Mental Health*, vol. 12, no. 1, pp. 1–4, Dec. 2018, doi: 10.1186/s13034-018-0214-7.
[34] W. R. DiPietro, "Political repression and government effectiveness," *Asian J. Social Sci. Stud.*, vol. 1, no. 1, p. 27, Feb. 2016, doi: 10.20849/ajsss.v1i1.14.
[35] L. Bode and E. K. Vraga, "In related news, that was wrong: The correction of misinformation through related stories functionality in social media," *J. Commun.*, vol. 65, no. 4, pp. 619–638, 2015, doi: 10.1111/jcom.12166.

NOOR AFIZA MAT RAZALI received the bachelor's degree in computer and information engineering, the Master of Science degree in computer science, and the Ph.D. degree in computer science from Japanese Universities. She is currently an Associate Professor with the Faculty of Defence Science and Technology, National Defence University of Malaysia. She also appointed as a Visiting Lecturer and a fellow at Management and Science University Malaysia. She is also a Professional Technologist, a recognition given by the Malaysia Board of Technologist. Prior to joining the university as an Academician, she was in the IT industry with nine years of working experience in multiple international corporations focusing on IT solution implementations, operation support, and project management, involving stakeholders in Malaysia, Japan, United States, and European Countries. Her research and expertise are in the areas of cybersecurity, national security, big data analytics, human–computer interaction, artificial intelligence and robotics, and blockchain technology.

NUR ATIQAH MALIZAN received the degree in computer science in 2020. She is currently pursuing the master's degree with the National Defence University Malaysia. She is also a Research Assistant with the National Defence University Malaysia. Her research interests include data mining, systems analysis, and big data analytics.

NOR ASIAKIN HASBULLAH received the Bachelor of Science degree (Hons.) in information technology from the National University of Malaysia, the Master of Science degree in information technology from the MARA University of Technology Malaysia, and the Ph.D. degree in IT and quantitative sciences (system sciences). She is currently a Senior Lecturer with the Department of Computer Science, National Defence University of Malaysia. She is also a Professional Technologist, recognition given by the Malaysian Board of Technologist. Her research interest and expertise are in the field of intersection between society and technology, such as information privacy, ethics in ICT, social informatics, and social network behavior.

MUSLIHAH WOOK was born in Batu Gajah, Perak. She received the bachelor's degree (Hons.) in information technology from Universiti Utara Malaysia, in 2001, the master's degree in computer science from Universiti Putra Malaysia, in 2004, and the Ph.D. degree in information science from Universiti Kebangsaan Malaysia, in 2017. Currently, she is working as a Senior Lecturer with the Department of Computer Science, Faculty of Defence Science and Technology, National Defence University of Malaysia. Her research interests include data mining applications in various domain, particularly in education, security, and defence. She has been a member of the International Association of Computer Science and Information Technology (IACSIT) and the Institute of Research Engineers and Doctors (IRED), since 2011 and 2013, respectively. Recently, she has been appointed as a Technical Reviewer of *Education and Information Technologies*—Springer's journal indexed by Scopus (Q2).

NORULZAHRAH MOHD ZAINUDDIN received the M.Sc. degree from Universiti Putra Malaysia and joined Military Academy of Malaysia, in 2002. She is currently a Lecturer with the Department of Computer Science, Faculty of Defence Science and Technology, National Defence University Malaysia (UPNM). She is also a member of the Informatics Intelligence Special Interest Group, UPNM. She has published a number of papers in international journals and conferences. Her main research interests include forensic computing, online social networks, and computer intelligence.

KHAIRUL KHALIL ISHAK received the bachelor's degree in engineering from Waseda University and the master's degree from Universiti Teknologi MARA Malaysia. He is a professional in information technology and engineering fields with experience in multiple multinational companies. His research interests include cyber security, big data analytics, artificial intelligence and robotics, and blockchain technology.

SUZAIMAH RAMLI was born in Temerloh Pahang. She received the bachelor's degree (Hons.) in information technology from Universiti Utara Malaysia, in 1997, the master's degree in computer science from Universiti Putra Malaysia, in 2001, and the Ph.D. degree in electrical, electronic, and system engineering from Universiti Kebangsaan Malaysia, in 2011. Currently, she is working as an Associate Professor with the Department of Computer Science, Faculty of Defence Science and Technology, National Defence University of Malaysia. She is also a member of the Malaysia Board of Technologist and Informatics Intelligence Special Interest Group, UPNM. Her research interests include image processing and artificial intelligence applications in various domain, particularly in education, security, and defence. She has published and presented most of her research findings to various international conferences and articles in many international journals specifically in her research niche.

SAZALI SUKARDI graduated from the Joint Services Command and Staff College, Shrivenham, U.K. He received the Bachelor of Science degree in civil engineering from the University of Hartford, CT, USA, the Post Graduate Diploma degree in system analysis from Universiti Teknologi MARA, the Master of Arts degree in defence studies from the King's College, University of London, and the Master of Science degree in information security from Royal Holloway, University of London. He is currently a Senior Vice President of Strategic Research, CyberSecurity Malaysia, an agency under the Ministry of Communication and Multimedia, Malaysia. Prior to joint CyberSecurity Malaysia, he has served at Malaysian Armed Forces for 22 years before retiring in 2009. He deals with various matters pertaining to cyber security strategic research and exploration of new cyber security initiatives and policies.