iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: III Month of publication: March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67445

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

Rohit Kumar Ojha[1], Panduga Mani Prasad Goud[2], Puligandla Udaya Sri[3], B. Sujani[4]
[1, 2, 3]*Student, Department of CSE(AI&ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India*
[4]*Assistant Professor, Department of CSE(AI&ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India*

Abstract: The internet provides a strong medium for the expression of opinions, emotions and ideas, through online media backed by smartphone use and high internet penetration. The majority of internet postings are text-based and may encompass people's emotional sentiments for a specific moment or feeling. Online sentiments or opinions should be monitored to identify any violent emotions evoked by citizens that can cause harm. Riots and warfare, for example, need to be tackled because there is a threat of undermining social stability and political security, which are very essential components of nations security. Mining opinions for national security is a pertinent research issue that needs to be developed further. Mechanisms and methods capable of mining opinions in the area of political security need to be greatly improved to yield optimal results. Academics have seen that there is also an intimate connection between emotion, sentiment and political security breaks. This research posits a new theory framework for forecasting political security threats based on a hybrid method: the integration of lexicon-based methods and machine learning in cyberspace. In the suggested framework, Decision Tree, Naive Bayes, and Support Vector Machine have been used as threat classifiers. To confirm our suggested framework, an experimental analysis is achieved. The performance of every technique employed in the experiments is presented. In this research, our suggested framework indicates that the hybrid Lexicon-based method with the RandomForest classifier achieved the highest performance score for political security threat prediction. Results show current research on opinion mining in threat prediction based on the political security domain.

Keywords: lexicon-based approach, Mining opinions, Decision Trees, Naive Bayes algorithm, SVM ALGO, political security threat

I. INTRODUCTION

The internet is now an integral part of national security in the modern era, with the US Intelligence Community identifying cyber threats as key concerns. Cyber threats are now being considered on a par like terrorism, indicating the dynamic nature of security threats. But protecting a country has become more complex with the overwhelming amount of data, the plethora of information on the internet, and the widespread dissemination of misinformation and false news. These as whole present a constant threat to national security. One most important element that this project touches on is the interplay between online opinions, sentiments, and security threats. It highlights the necessity of identifying and responding quickly to threats arising through online feelings and sentiments. As much as the apparent linkage between online expressed emotions and security threat exists, to date, there is significant lack a full-range assessment framework within the idea of nations security. This project seeks to fill this knowledge gap by innovating a new method for forecasting political threats that are tied to online emotions. In acknowledgment of the pivotal importance of emotions in framing public opinion and potentially affecting security dynamics, the project unites cutting-edge word analysis methods with machine learning models. By making use of actual news data, this hybrid system aims to bridge current knowledge gaps and offer practical insights into prospective security threats posed by online emotions. Through the convergence of analytical approaches with actual data, it would equip authorities with the ability to anticipate and resolve upcoming security issues, thus promoting political security and national safety in the contemporary era.

Under the current systems, the utilization of sentiments, or opinions, expressed in a piece of text has the potential to instill bad feelings or emotions like anger or fear that have a potential to cause events compromising national security. Because data communicated in cyberspace is often laced with emotions that could hold national security vulnerabilities (to each aspect of national security), Real Time detection of disruptive emotions contributes a significant role in assisting authorities in controlling the situation early.

Different gaps, methods and application domains that are centered on current opinion mining techniques (like the lexicon-based method and machine learning methods) can be employed to identify the current sentiments inherent in sentences across a number of domains, as explained in. The evaluation and framework analysis on emotions and their measurements in the context of national security is absent. Research work on opinion mining in the area of national security has not yet been investigated fully, but it can identify many threats and support the protection of a nation.

II. RELATED WORK

Over the past few years, using machine learning and lexicon-based methods for forecasting political security threats has become increasingly popular. Rule-based and keyword-matching methods have been used for security threat analysis but tend to be inadequate when dealing with contextual awareness and dynamic political language. To overcome this limitation, several hybrid lexicon-based and machine and learning models have been investigated to improve threat accuracy and resilience.

1) Statement for the record: Global Threat Assessment by the U.S. Intelligence Community

J. R. Clapper,2015

This one sentence offers considerable information regarding many state and nonstate players, cutting across political, economic, and military advancements and transnational trends, all of which are our country's strategic and tactical environment. While I think counterterrorism, counterproliferation, cybersecurity, and counterintelligence are most immediately on the minds of our security community, it is well nigh impossible to prioritize—based on ultimate significance—the many, possible threats to US national security. The United States no longer confronts—one such as during the Cold War—one overarching threat. Instead, it is the interrelatedness and multiplicity of possible threats—and the parties responsible for them—that represent our greatest challenge

2) Opinion Mining Techniques for National Security Analysis domain applications research and challenges opportunities

N. A. M. Razali et al.,2021.

Opinion mining, also known as sentiment analysis field. It retrieves individuals' thoughts, such as judgments, attitudes, and feelings toward people, subjects, and programs. This process is technically difficult but extremely helpful. With the sudden rise of the digital media in the world, like blogs and social media networks, people and administrations are using public opinion more and more for decision-making. In the past years, there was a lot of research regarding mining people's sentiments from text in cyberspace with the help of opinion mining. Researchers have utilized various opinion mining methods, such as machine learning and lexicon-based methods to analyze and categorize individuals' sentiments according to a text and discuss the gap that exists. Therefore, it presents research opportunities for other researchers to study and suggest better methodologies and new domains and applications to bridge the gap.

3) Sentimental analysis of the methods and approach: Survey

S. Dorle,2017

Today, social media is a great source of business decision support and Data Analytics is also used by too many industries and organizations to make a more improved Business decision. Using analytics to the data the enterprises make a tremendous difference in their planning and decision-making approach. Sentiment analysis or opinion mining is a great contribution in our day-to-day decision-making process. These choices can vary from buying a product like a mobile phone to watching the movie to investing. All choices will have a significant effect on day-to-day life. Sentiment Analytics or Opinion analysis is done to find out the opinion of people. It is being used as a Lexicon approach Based on Machine and Learning based approach. Few approaches are still inefficient in extracting the sentiment features of the provided content of text. Naive Baye algo, Supported Vector Machines algo are the machine learning algorithms applied in sentiment analysis that only has limited sentiment classification category between positive and negative. Although the development of sentiment Analytics technique there are a number of problems still to be observed and render the analysis not precisely and effectively. Therefore, the paper is illustrating the survey of different sentiment Analytics Methodologies and approaches. It will be beneficial to gain transparent knowledge on sentiment analysis methodology.

4) Opinion mining on newspaper quotations

A. Balahur, R. Steinberger, E. Van Der Goot, B. Pouliquen and M. Kabadjov, 2009.

Opinion mining is the process of gathering from a collection of documents the opinions voiced by a source about a given target. The article here gives a comparative overview of the approaches and tools that may be used to mine opinions from quotations (reported speech) found in newspaper reports. We demonstrate the challenge of this task, encouraged by the occurrence of various possible targets and the vast range of affected phenomena quotes hold. We assess our methods using annotated quotations drawn from news supplied by the EMM news gathering engine. We find that a generic opinion mining system necessitates both large lexicon use, as well as specialized training and test data.

5) Analysis of Opinion Using a Combination of Lexicon Method and Multinomial Bayes Classifier

G. Isabelle, W. Maharani and I. Asror,2019.

Opinion mining refers to the examination of the opinion by examining the sentiment, behavior, or emotion encompassed in a product. Lexicon-based and supervised learning are some methods used for mining opinions. Lexicon-based methods have low recall but good accuracy with long training time. Hence this paper will combine lexicon approach with one of supervised learning approaches, i.e. Multinomial Naïve Bayes for English language and identify opinion based on sentiment class, i.e., positive and negative. Feature extraction used in this research are unigram, POS-Tagging, and score-based feature on lexicon. The system output is the polarity of the document and the performance will be measured by Precision, Recall, and F-measure. By using opinion mining with combining lexicon-based methods and Multinomial Naive Bayes, this research achieved accuracy 0.637.

III. PROPOSED METHODOLOGY

We introduced a new Hybrid Lexicon and machine learning-based algorithms in the proposed system. Both algorithms such as NRC lexicon and Machine learning will be combined to create Hybrid algorithm where NRC lexicon will be utilized to compute emotions from the news or public opinions and then both news and emotion labels will be kept to machine and learning model to train a model and this model can be used on any news to predict Threat or No Threat Label. A most important element addressed by this project is the connectivity between online opinion, sentiment, and security threat. It places emphasis on being able to detect and respond in a timely fashion to new threats ascertained through online sentiments and opinions. In spite of the clear connectivity between online-expressed emotions and security threats, there is now a significant lack of an adequate assessment framework among national security academia. This project seeks to close this gap by innovating a novel methodology for forecasting political dangers associated with online emotions. Given the pivotal importance of emotions in framing public debate and potentially affecting security dynamics, the project combines sophisticated word analysis methods with machine learning approaches. Utilizing actual news data, the hybrid framework attempts to bridge any current knowledge divides and offer concrete insights into existing security threats presented by online feelings.

Dataset Used : NRCLexicon is an MIT-approved pypi project by Mark M. Bailey which predicts the sentiments and emotion of given text. The package contains approximately 27,000 words and is on the National Research Council Canada (NRC) affect lexicon and the NLTK library's WordNet synonym sets.

In the proposed system, introducing novel Hybrid Lexicon and machine learning based algorithms. Both algorithms like NRC lexicon and Machine learning will get combined to form Hybrid algorithm where NRC lexicon will be used to calculate emotions from the news or public opinions and then both news and emotion labels will be input to machine and learning algo for training the model and this model can be applied on any news to predict Threat or No Threat Label.

A. Modules Description

- Step 1: Upload Dataset By using this module, we can upload a dataset for training the algorithms.
- Step 2: Dataset preprocessing: By using this module, we can split pre trained dataset for training and testing
- Step 3: Run NaiveBayes Algorithm: By using this module, we can train NaiveBayes Algorithm for predicting political threat and getting accuracy as 89%
- Step 4: Run SVM Algorithm: By using this module, we can train SVM Algorithm for predicting political threat and getting accuracy as 89.50%
- Step 5: Run Decision Tree Algorithm: By using this module, we can train and getting accuracy as 92%
- Step 6: Run Random Forest Algorithm: By using this module, we can train random forest classifier for the political threat and getting accuracy as 93%
- Step 7: Comparison Graph: By using this module, we then compare all the algorithms with accuracy, precision, recall, f-score
- Step 8: Upload Test Data: By using this module, we can upload test data for predicting whether data belongs to threat or no threat.

B. Architecture

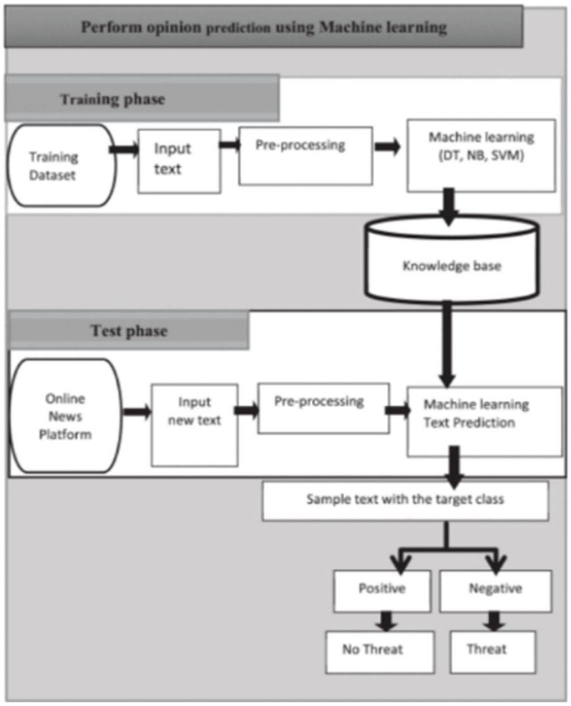


Fig 1. Architecture

IV. EXPERIMENTAL RESULTS

Experimental assessment of the suggested politics secure threatful prediction was carried out with the application of several machine and learning models such as Naive Bayes classifier, (SVM), Decision Tree, and Random Forest. The main outcomes of the experiments are provided below:

A. Naïve Bayes Algorithm

Accuracy: 89%
Naïve Bayes classifier was trained on political security threat data and exhibited robust performance in text classification. It was less accurate than tree-based models because it assumed feature independence.

B. Support Vector Machine (SVM) Algorithm

Accuracy: 89.50%
SVM model performed little than better Naïve Bayes, exhibiting robust generalization in classifying political security threats. It effectively classified threat and non-threat instances with hyperplane-based classification.

C. Decision Tree Algorithm

Accuracy: 92%
The Decision Trees model also performed better than Naïve Bayes and SVM, using hierarchical decision-making to classify security threats efficiently.

D. Random Forest Algorithm

Accuracy: 93%
The Random Forest model yielded the highest accuracy, showcasing strong prediction capabilities by using several decision trees to minimize variance and enhance performance.

E. Comparison of Performance Metrics

A comparison graph was plotted to represent the accuracy, precision, recall, and F-score of all the classifiers, where it was verified that the Random Forest provided the top performance.

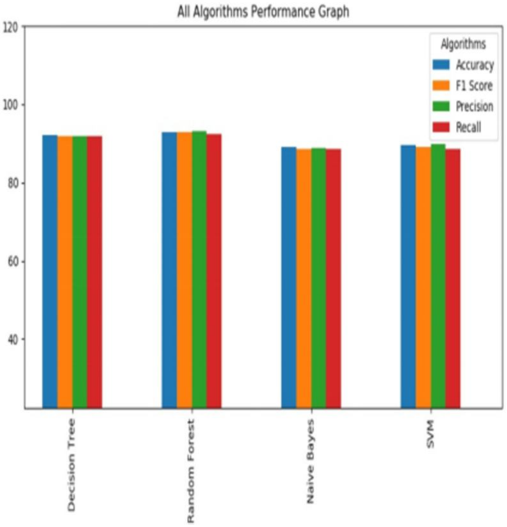
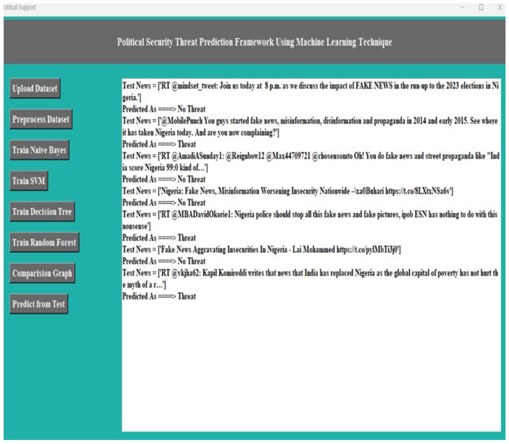


Fig 2. Performance Comparison graph

The system efficiently recognizes potential security risks by processing news stories and public sentiments through a hybrid lexicon-based and machine and learning method. The system incorporates NRC Lexicon for sentiment analytics and machine and learning algorithms (Naive Bayes, SVM, Decision Tree, and Random Forest) for classification. The Random Forest model performed the top with highest accuracy (93%), showing the top performance in differentiating between "Threat" and "No Threat" classes. This app offers an important real-time threat detection tool that allows the authorities to be proactive in safeguarding national security.



```
Test News = ['RT @mindset_tweet: Join us today at  8 p.m. as we discuss the impact of FAKE NEWS in the run-up to the 2023 elections in Ni
geria.']
Predicted As ====> No Threat
Test News = ['@MobilePunch You guys started fake news, misinformation, disinformation and propaganda in 2014 and early 2015. See where
it has taken Nigeria today. And are you now complaining?']
Predicted As ====> Threat
Test News = ['RT @AmadiASunday1: @Reignbow12 @Max44709721 @chosensomto Oh! You do fake news and street propaganda like "Ind
ia score Nigeria 99:0 kind of...']
Predicted As ====> No Threat
Test News = ['Nigeria: Fake News, Misinformation Worsening Insecurity Nationwide -\xa0Buhari https://t.co/8LXtxNSa6v']
Predicted As ====> No Threat
Test News = ['RT @MBADavidOkorie1: Nigeria police should stop all this fake news and fake pictures, ipob ESN has nothing to do with this
nonsense']
Predicted As ====> Threat
Test News = ['Fake News Aggravating Insecurities In Nigeria - Lai Mohammed https://t.co/pylMbTi3j0']
Predicted As ====> No Threat
Test News = ['RT @vkjha62: Kapil Komireddi writes that news that India has replaced Nigeria as the global capital of poverty has not hurt th
e myth of a r...']
Predicted As ====> Threat
```

Fig 3. News articles and public opinions analysis

V. DISCUSSION

The Political Security Threat Prediction System is intended to detect potential security threats through the analysis on online news stories and public sentiments. The system includes and integrates Natural Language Processing (NLP) methods with machine learning algorithms to categorize text data into "Threat" or "No Threat" classes. Through the use of a hybrid lexicon-based and machine learning method, the system provides a more complete and precise evaluation of potential threats.

The central part of system is lexicon-based sentiment analysis, whereby the NRCLex library extracts emotional features of text. It enables the system to determine a negative and positive sentiment score based on the possibility of a text being related to a political threat. If there are more negative sentiments than positive sentiments, the message is tagged as a "Threat"; else, it's tagged as "No Threat." The sentiment is augmented with TF-IDF vectorization, where the text is transformed into a machine learning processing-friendly structured form.

For classifying, the system uses a combination of ML algorithms, such as Naive Bayes, (SVM), Decision Tree, and Random Forest. Every model is trained using a labeled dataset to improve its predictive power. The Random Forest model showed the best accuracy at 93%, which made it the best classifier among the algorithms that were tested. The performance of the system was measured using accuracy, precision, recall, and F1-score to ensure a valid and balanced evaluation of political security threats.

Through the integration of sentiment analysis and machine learning, the system presents a strong and scalable solution for real-time identification of potential security threats. This system presents a useful tool for law enforcement organizations, government agencies, and security analysts, allowing proactive decision-making as well as enhanced national security measures. The integration of hybrid lexicon-based techniques and AI-driven classification models ensures that the system remains adaptive and efficient in analyzing dynamic online content, making it a significant advancement in political security threat prediction.

VI. CONCLUSION FOR FUTURE WORK

In this research, we presented a hybrid approach that combines the lexicon-based method with machine learning to forecast political security risks in cyberspace. Given the increasing impact of social media and online sites, bad actors frequently employ such venues to disseminate extremist ideas, disinformation, and threats to national security. Our method blends sentiment analysis and machine learning classifiers—Decision Tree, Naïve Bayes, and Support Vector Machine (SVM)—to examine and categorize online discussion. By utilizing lexicon-based sentiment analysis, we managed to capture emotional sentiments and views from text-based information, which enabled our classifiers to efficiently pinpoint possible security threats.

Experimental findings showed that though classical classifiers such as Decision Tree and Naïve Bayes did the job, the Random Forest classifier, when combined with the lexicon-based method, was most accurate in security threat prediction. This reiterates the strength of ensemble learning methods in enhancing classification accuracy. Our results identify a high correspondence between online opinions and political security threats, reiterating the need for opinion mining tasks in nation security in the real time environment. Through the combination of structured and unstructured sources of data, our framework improves digital security monitoring and offers a more dependable way of early threat discovery.

This study adds to the larger body of cyber threat intelligence, providing policymakers, security organizations, and law enforcement with important insights to actively counter new risks. The framework suggested can be applied to track global political events, cyber warfare operations, and disinformation campaigns, sending real-time alerts regarding possible security threats. Further increasing the dataset with multilingual sources and regional dialects can enhance flexibility in varying geopolitical environments, thus making the system more robust for global security uses

REFERENCES

[1] J. R. Clapper, “Statement for the record: Worldwide threatful task of the US intelligence community,” 2015.
[2] P. Barnaghi, J. G. Breslin, I. D. A. B. Park and L. Dangan, 2016, “Opinion mining task and sentimental polarity on Twitter and correlation between events and sentiment.”
[3] A. Balahur, R. Steinberger, E. Van Der Goot, B. Pouliquen and M. Kabadjov, 2009, “Opinion mining on newspaper quotations.”
[4] G. Clark, W. Maharani and I. Asror, 2019, “Analysis of the opinions of mining and using the combining lexicon methods and multinomal Naïve Bayes.”
[5] W. Zhang, W. Gan and B. Jiang, 2014, “Machine learning and lexicon-based methods for sentiment classification: A survey.”
[6] V. Vaitheeswaran and D. L. Arockiam, 2016, “Combining nrc and machine learning methods to enhance the accurately of sentiment analysis on big data.”
[7] Z. Shaikh and A. Fatemah Meghji, 2019, “Sentiment analysis of news articles: A lexicon-based approach.”
[8] B. Ding, B. Liu and P. S. Yu, 2008, “A holistic lexicon approach to opinion mining.”
[9] S. Razali, 2021, “Opinion mining task for nation security: Techniques, domain applications, challenges, and research opportunities.”
[10] J. Dorle, 2017, “Sentiment analysis methods and approach: A survey.”
[11] R. Zgheib and A. M. Barbar, 2017, “A study using svm to classify the sentiments of tweets.”
[12] D. Stein, E. Van Der Goot, and P. Pouliquen, 2018, “Analyzing political security threats using machine learning-based sentiment classification.”