

ESTABLISHMENT OF FRAMEWORK FOR PREDICTION OF POLITICAL THREATS USING HYBRID TECHNIQUES

Mr.B. SUMAN¹, INDHU CHENCHETI², ABHINAYA GOLLA³, LAVANYA
MADDILETI⁴

¹Assistant Professor, Department of Computer Science and Engineering, Sridevi Women's
Engineering College, Hyderabad, India

Email: swecsuman@gmail.com

^{2,3,4}B.Tech Student, Department of Computer Science and Engineering, Sridevi Women's
Engineering College, Hyderabad, India

Abstract: The internet offers a powerful medium for expressing opinions, emotions and ideas, using online platforms supported by Smartphone usage and high internet penetration. Monitoring online sentiments or opinions are important for detecting any excessive emotions triggered by citizens who can lead to unintended consequences and threats to 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. We propose a new theoretical framework for predicting political security threats using a hybrid technique: the combination of lexicon-based approach and machine learning in cyberspace. The proposed framework uses Decision Tree, Naive Bayes, and Support Vector Machine as threat classifiers. To validate our proposed framework, an experimental analysis is accomplished. The framework reveals that the hybrid Lexicon-based approach with the Decision Tree classifier recorded the highest performance score for predicting political security threats. Natural Language Processing (NLP) can be applied in opinion mining.

Index Terms - Internet, Sentiments, Online Platforms, Smartphone Usage, National Security, Opinion Mining, Theoretical Framework, Hybrid Technique, Lexicon-based Approach, Machine Learning.

Keywords: Sentiment classification, Neural Language Model, Machine Learning, Natural Language Processing, Textual Data, Opinion Mining, Topic Augmentation.

1. 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. 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. 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. Various gaps, techniques and domain applications that focus on existing opinion mining methods can be used to determine the existing sentiments embedded in sentences throughout several domains.

In recent years, the global landscape has witnessed a heightened focus on national security, prompting a comprehensive exploration of emerging threats. J. R. Clapper, in a pivotal statement for the record to the US intelligence community in 2015 [1], emphasized the imperative of understanding worldwide threats. As technology continues to evolve, so does the need for innovative approaches to safeguarding nations.

The burgeoning field of opinion mining, as discussed by Razali et al. [2], has emerged as a promising avenue for bolstering national security efforts. Opinion mining, also known as sentiment analysis, involves extracting and evaluating opinions from diverse sources, such as social media, news articles, and public discourse. This paper delves into the techniques, applications, challenges, and research opportunities associated with opinion mining for national security [2]. Sentiment analysis methods and approaches have undergone significant scrutiny [3], with research exploring avenues such as sentiment analysis on newspaper quotations [4], sentiment classification on Twitter [6], and sentiment analysis in various languages, including Malayalam [10].

A survey by Ravi and Ravi [7] comprehensively outlines tasks, approaches, and applications in opinion mining and sentiment analysis, shedding light on the multifaceted nature of this domain. Combining lexicon-based methods with machine learning approaches has been a recurring theme in

enhancing the accuracy of sentiment analysis [8] and classifying sentiments in diverse contexts, such as tweets [14]. The synthesis of lexicon and machine learning methods represents a contemporary strategy to navigate the intricacies of sentiment analysis on big data [15]. This introduction sets the stage for a detailed exploration of the methodologies, challenges, and advancements in opinion mining for national security, drawing on a rich tapestry of research contributions [2-15].

2. LITERATURE REVIEW

The literature surveyed encompasses a wide range of studies on opinion mining and sentiment analysis, with a particular focus on applications in national security, techniques employed, challenges encountered, and avenues for further research. The provided citations form a comprehensive overview of the field, spanning from global threat assessments to specific methods and approaches used in sentiment analysis. In the realm of national security, Clapper's statement [1] highlights the importance of intelligence assessments in understanding global threats. To enhance national security, Razali et al. [2] delve into opinion mining, discussing techniques and applications. This provides a foundation for understanding how sentiment analysis can be applied to extract valuable insights for security purposes.

Dorle [3] offers a survey of sentiment analysis methods and approaches, providing a broader understanding of the landscape. Balahur et al. [4] specifically focus on opinion mining applied to newspaper quotations, showcasing the diverse sources of data that can be analyzed for sentiment. Seerat's work [5] delves into the issues and challenges associated with opinion mining, shedding

light on potential hurdles in implementing these techniques. The exploration of sentiment analysis on social media platforms is evident in the works of Barnaghi et al. [6], Ravi and Ravi [7], and Zhang et al. [9].

Barnaghi et al. specifically examine sentiment polarity on Twitter and its correlation with events, demonstrating the relevance of social media in gauging public sentiment. Ravi and Ravi's survey [7] provides a comprehensive overview of sentiment analysis tasks, approaches, and applications, emphasizing the significance of sentiment analysis across various domains. Lexicon-based approaches are extensively discussed in the literature. Ding et al. [12] propose a holistic lexicon-based approach to opinion mining, highlighting the importance of comprehensive lexicons. Gheib and Barbar [14] and Vaitheeswaran and Arockiam [15] explore the integration of lexicon and machine learning methods to enhance sentiment analysis accuracy, indicating the evolving strategies in this field.

Additionally, language-specific sentiment analysis is addressed by Ashna and Sunny [10] in their study on sentiment analysis in the Malayalam language, emphasizing the need for language-specific approaches in diverse linguistic landscapes. Similarly, Taj et al. [11] apply a lexicon-based approach to sentiment analysis of news articles, showcasing the versatility of sentiment analysis techniques across different content types.

Saberi and Saad [13] provide a comprehensive review of sentiment analysis or opinion mining, summarizing various methodologies and their applications. This review serves as a valuable resource for researchers and practitioners looking to

understand the broader landscape of sentiment analysis. In conclusion, the literature survey reveals a rich tapestry of research in the field of sentiment analysis and opinion mining, with applications ranging from national security assessments to social media and news articles. The integration of lexicon-based approaches and machine learning methods, along with language-specific considerations, showcases the diverse strategies employed in understanding and analyzing sentiments. As technology evolves, these studies lay the groundwork for future research endeavors in sentiment analysis.

3. METHODOLOGY

In previous work they introduced method to analyze the sentiment polarity of Twitter posts using a machine learning method for text categorization called Bayesian Logistic Regression (BLR) Classification. Their aim is to determine whether a tweet expresses a positive or negative sentiment towards a given topic. They also focus on building a trained model to accomplish this task and to look for correlations between Twitter sentiment and major events, using the FIFA World Cup 2014 as a case study. They used Twitter Streaming API and official World Cup hash tags to mine, filter, and process tweets in order to analyze the reflection of public sentiment towards unexpected events. In another research they introduced a lexicon-based approach for sentiment analysis of news articles. They have performed experiments on a BBC news dataset to validate the applicability of their approach.

Drawbacks:

1. The existing work uses Bayesian Logistic Regression for text categorization, but it

may not fully capture the emotional aspect of the sentiments expressed in tweets. the lack of emotional analysis might limit the depth and accuracy of the results.

2. This approach may lead to biased or incomplete data representation, as not all tweets related to the event might use the designated hashtags.
3. The existing work is specifically designed for sentiment analysis on Twitter. While Twitter is a popular platform for real-time updates and public opinions, its 280-character limit and informal language might limit the depth and context of the sentiments expressed.
4. The existing work, which is using a lexicon-based approach for sentiment analysis, is its limited capability to handle context and sarcasm effectively.

In this study, We 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 which 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.

Benefits:

1. Our work gathers data from online news platforms, which might provide a more comprehensive and diverse dataset.
2. Our work focuses on online news platforms, which generally provide more extensive and detailed information.
3. This hybrid approach is likely to enhance the accuracy and robustness of the predictions compared to the single-method approach.
4. Our hybrid technique, which combines a lexicon-based approach with machine learning, might offer improved performance, as machine learning algorithms can better adapt to different contexts and learn from data to capture the complexity of language used in online news articles related to political security.

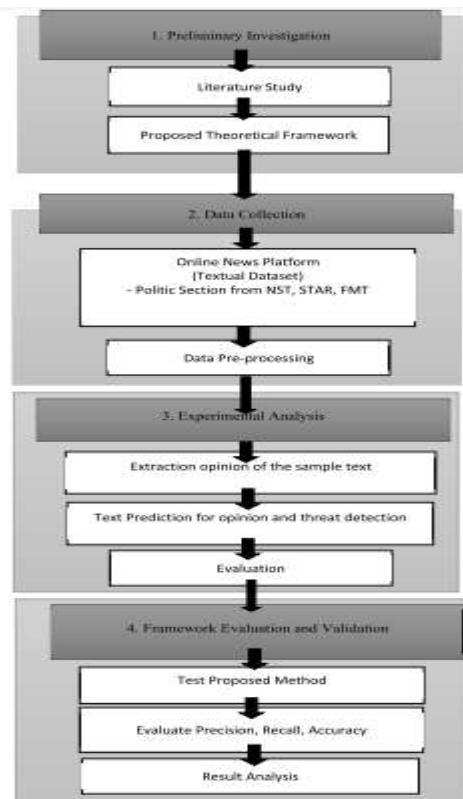


Fig 1 System Architecture

Modules:

The modules are:

- Data loading: using this module we are going to import the dataset.
- Data Preprocessing: using this module we will explore the data.
- Splitting data into train & test: using this module data will be divided into train & test
- Model generation: Building the model - Naive Bayes -SVM -Decision Tree -Voting Classifier (AB + RF) -Stacking Classifier. Algorithms accuracy calculated

- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4. IMPLEMENTATION

Naive Bayes –It is a classification technique based on Bayes' Theorem with an independence assumption among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

SVM –SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as margin. And the goal of SVM is to maximize this margin. The hyperplane with maximum margin is called the optimal hyperplane.

Decision Tree –A decision tree algorithm is a machine learning algorithm that uses a decision tree to make predictions. It follows a tree-like model of decisions and their possible consequences. The algorithm works by recursively splitting the data into subsets based on the most significant feature at each node of the tree.

Voting Classifier (AB + RF) – A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output.

Stacking Classifier- Stacking is an ensemble machine learning algorithm that learns how to best combine the predictions from multiple well-performing machine learning models. The scikit-learn library provides a standard implementation of the stacking ensemble in Python.

5. EXPERIMENTAL RESULTS

Dataset Description:

The "Indian Politics Tweets & Reactions" dataset contains a collection of public tweets and their associated reactions related to Indian politics. The dataset provides insights into public opinion and sentiment on various political issues and events in India, as expressed on Twitter. The dataset includes a diverse range of tweets from users across the country, covering topics such as political campaigns, government policies, elections, political parties, leaders, and controversies. The reactions to each tweet are also included, such as likes, retweets, and comments, allowing for a comprehensive analysis of the engagement and impact of each tweet. Researchers and analysts can use this dataset to study the sentiment and trends of public opinion on Indian politics and to gain insights into the effectiveness of political campaigns and messaging on social media.



Fig 2 Code Folder



Fig 3 Anaconda prompt

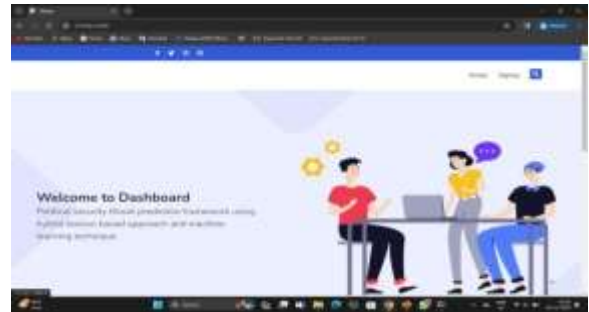


Fig 4Dashboard

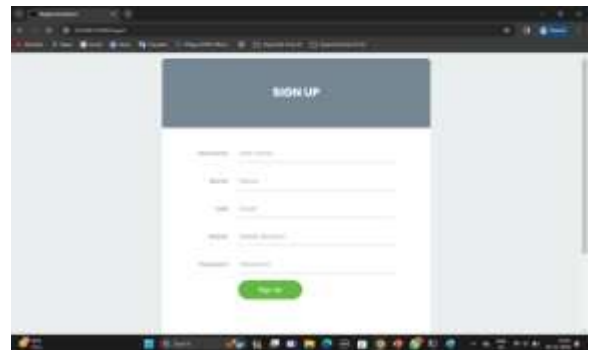


Fig 5SignUp page



Fig 6Signin Page

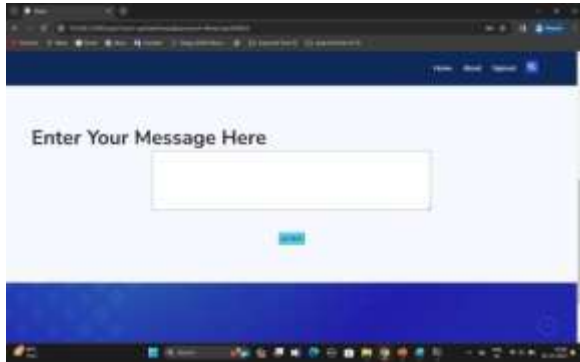


Fig 7 Message box

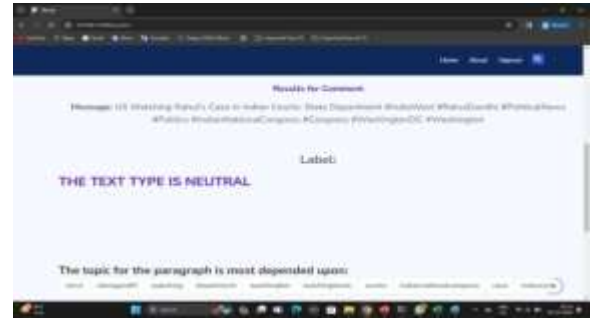


Fig 11 Sample 2 Result

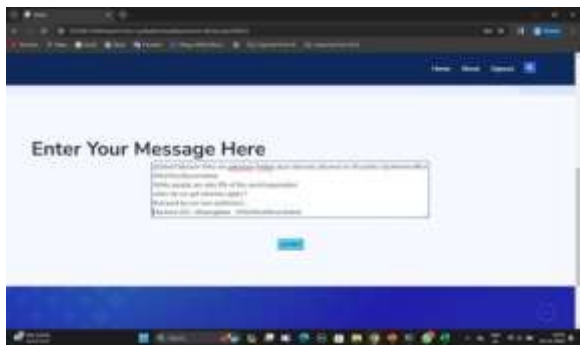


Fig 8 Sample Tweet 1

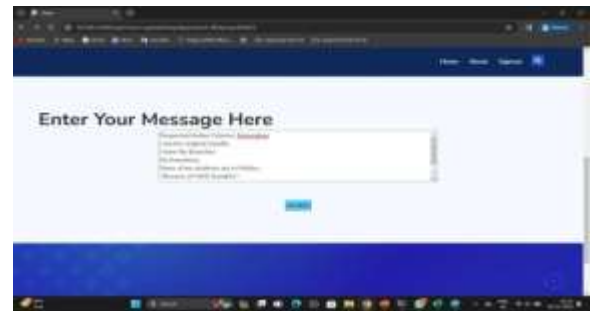


Fig 12 Sample Tweet 3



Fig 9 Sample 1 Result



Fig 13 Sample Result 3



Fig 10 Sample Tweet 2

6. CONCLUSION

Our proposed framework for predicting political security threats using a hybrid approach of lexicon-based analysis and machine learning techniques are designed to analyze people’s opinions on the national security domain, with a specific focus on the political security element. We 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. We successfully demonstrated the relationship between emotions, opinions, sentiment, and political security threats in cyberspace. We 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. We 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.

FUTURE WORK:

The future enhancement of the "Political Security Threat Prediction Framework" involves a multi faceted approach to bolster its effectiveness and adaptability. Firstly, integrating advanced semantic analysis techniques will enhance the framework's ability to decipher the nuanced context within political texts, refining threat predictions. Real-time data streaming from diverse sources, including social media and news outlets, will ensure that the model remains up-to-date, capturing the latest developments. Multi-modal data fusion, incorporating image and video processing alongside textual analysis, promises a more comprehensive understanding of potential threats. Deep learning architectures, such as recurrent neural networks (RNNs) or transformers could be explored to capture intricate patterns in political discourse, elevating

prediction accuracy. The framework's lexicon should dynamically evolve to stay relevant, and a user feedback mechanism can be integrated for continuous improvement based on real-world insights. Geospatial analysis will provide a nuanced understanding of regional variations in political threats. Improving explainability and interpretability, incorporating cross-domain analysis capabilities, and fostering collaborative intelligence among different entities will further strengthen the framework's capabilities in predicting and addressing political security threats.

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.