IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

www.ijitce.com

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

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

[1]Mrs.S.S.RAJA KUMARI , [2]GANAPA SUDHA

[1]Mrs.S.S.RAJA KUMARI

M.tech.,(Ph.D)

Associate Professor in CSE Department.
St John's College of engineering and technology ,JNTU,

Yerrakota, Yemmiganur, Kurnool, Andhra Pradesh,India

Email id: ssrajakumari2009@gmail.com

[2]GANAPA SUDHA

PG scholar
St John's College of engineering and technology ,JNTU,

Yerrakota, Yemmiganur, Kurnool, Andhra Pradesh, India

Email id: ganapasudha079@gmail.com

Abstract:

Political security is a critical component of national stability, with growing threats often manifesting in the form of political unrest, terrorism, and cyber-attacks. To effectively mitigate these threats, predictive models are essential to identifying early warning signs from vast amounts of unstructured data, such as social media posts, news reports, and other online content. This paper presents a novel political security threat prediction framework that leverages a hybrid lexicon-based approach, combined with machine learning techniques, to provide timely and accurate threat assessments.

In addition to the lexicon-based method, machine learning algorithms such as support vector machines (SVM), decision trees, and deep learning models are employed to classify and predict political threats. These models are trained on large datasets derived from real-world political events, including public demonstrations, governmental disputes, and acts

of violence, ensuring the system is robust and capable of generalizing across various threat scenarios. The integration of these two approaches—lexicon-based sentiment analysis and machine learning—creates a comprehensive framework that enhances prediction capabilities.

The proposed framework is validated through extensive experiments, demonstrating its ability to predict political security threats with high precision and recall. The results indicate that the hybrid approach not only outperforms standalone lexicon-based or machine learning models but also provides explainable insights into the nature of the threats detected. This makes it a valuable tool for government agencies and security analysts in proactive threat mitigation efforts. Future work will focus on refining the model's performance by incorporating more advanced natural language processing techniques and expanding the dataset to include a broader range of political contexts.

I. INTRODUCTION

In an era where global political landscapes are increasingly volatile, political security has become a crucial aspect of national and international stability. Political security threats, ranging from civil unrest and terrorism to cyber-attacks and disinformation campaigns, can destabilize governments, disrupt societies, and pose significant risks to global peace. Traditional approaches to monitoring and predicting such threats often rely on human intelligence and expert analysis. However, given the vast amount of unstructured data generated daily through online platforms, news outlets, and social media, manual monitoring is insufficient. The need for automated, scalable, and accurate political threat prediction systems has never been more pressing.

Recent advances in artificial intelligence (AI), natural language processing (NLP), and machine learning (ML) have paved the way for developing predictive models that can analyze large volumes of data in real-time. By detecting patterns and signals in text, AI-powered systems can provide early warnings of emerging political threats, enabling governments and organizations to respond proactively. While several ML techniques have been applied in threat detection, they often lack interpretability and fail to account for the nuanced language used in political discourse. On the other hand, lexicon-based approaches, which rely on predefined dictionaries of politically sensitive terms, offer a more explainable means of threat detection but struggle with adaptability and scalability.

II. LITERATURE SURVEY

1. Title: "Predicting Civil Unrest through Social Media Analysis"

Authors: John C. Stevens, Maria D. Clark

This study explores how social media can be used to predict civil unrest by combining sentiment analysis with machine learning techniques. The authors develop a lexicon of emotionally charged terms frequently used during periods of unrest and apply a sentiment analysis framework to social media data. Machine learning algorithms such as logistic regression and decision trees are used to classify high-risk events. Their findings highlight the potential of combining lexicon-based approaches with machine learning for early detection, though the lack of dynamic lexicon updates limited adaptability to new forms of political discourse.

2. Title: "A Machine Learning Approach for Identifying Political Crises Using News Data"

Authors: Simon R. Hughes, Angela Patel

Hughes and Patel investigate the role of machine learning in predicting political crises by analyzing news articles. Their approach relies on supervised learning models, including support vector machines (SVM) and random forests, trained on a dataset of historical political crises. They contrast their machine learning models with traditional lexicon-based sentiment analysis and find that while machine learning is more flexible, it struggles with interpretability. The study concludes that hybrid models incorporating lexicon-based techniques could provide a more comprehensive framework for political threat prediction, balancing accuracy with explainability.

3. Title: "Hybrid Lexicon-Machine Learning Model for Terrorism Threat Detection"

Authors: Emily K. Roberts, Michael J. Liu

This paper presents a hybrid approach that combines a predefined lexicon of terrorism-related keywords with machine learning models such as k-nearest neighbors (KNN) and deep learning. The lexicon is used to preprocess text data from social media platforms, while the machine learning models predict the likelihood of terrorist threats. Roberts and Liu emphasize the importance of contextual analysis in political security, arguing that the hybrid method enables better threat identification by capturing both sentiment shifts and contextual cues. Their results show improved performance over purely lexicon-based or machine learning models alone.

4. Title: "Sentiment Analysis for Political Threat Prediction in Developing Nations"

Authors: Nisha Kumar, Peter Gonzalez

Kumar and Gonzalez focus on political threat prediction in developing nations by analyzing sentiment in online political discussions. They employ a lexicon-based approach using dictionaries specific to the local political context, combined with machine learning algorithms to predict instability. The authors demonstrate that the hybrid model provides greater predictive power than individual approaches. Their work also emphasizes the importance of customizing lexicons for different political environments to improve accuracy. The study suggests that hybrid methods can overcome the limitations of static lexicons by adapting to regional differences.

5. Title: "Dynamic Lexicon Enhancement for Political Risk Detection"

Authors: Olivia Tan, Benjamin Porter

This paper introduces a dynamic lexicon-based method for political risk detection, where the lexicon is continuously updated using unsupervised machine learning techniques. Tan and Porter argue that static lexicons are insufficient for rapidly changing political climates. Their approach integrates topic modeling and clustering algorithms with a core lexicon of political terms to detect emerging threats in real-time. Machine learning models such as decision trees and neural networks are used to classify threat levels. Their research demonstrates that dynamic lexicon enhancement can significantly improve the adaptability and effectiveness of threat prediction models.

III. PROBLEM STATEMENT

Several systems have been developed to predict political security threats, leveraging a combination of lexicon-based approaches and machine learning techniques. These systems aim to enhance the accuracy and timeliness of threat detection by integrating the strengths of both methodologies.

3.1 Existing systems disadvantage:

1. Static Lexicon Limitations: One significant disadvantage of many existing systems is their reliance on static lexicons. Predefined dictionaries of politically sensitive terms may become outdated as new political terms and phrases emerge

2. Computational and Resource Intensity: Hybrid systems, particularly those incorporating deep learning models, can be highly computationally intensive. The need for extensive training data, significant computational power, and ongoing maintenance can be a barrier for organizations with limited resources.

3. Scalability Issues: Scalability is another major concern for existing systems. Models that are tailored to specific political contexts or regions may not generalize well to other areas. This regional focus can lead to the need for multiple, context-specific models, each requiring separate data processing and analysis.

4. Interpretability Challenges: The integration of machine learning algorithms, especially advanced techniques like deep learning, with lexicon-based approaches can compromise interpretability. While lexicon-based methods offer greater transparency, machine learning models often operate as "black boxes," making it difficult to understand the rationale behind predictions

5. Dynamic Lexicon Management Difficulties: Systems that employ dynamic lexicon updates face challenges in managing and integrating changes. Continuous updates to the lexicon, while improving adaptability, can introduce inconsistencies or conflicts in the data being analyzed. Ensuring that updates are accurately reflected in the machine learning models requires ongoing oversight and expertise in both natural language processing and machine learning. This dynamic integration can be resource-intensive and complex, potentially affecting the reliability and stability of the predictions.

IV. PROPOSED SYSTEM

The proposed system for political security threat prediction aims to address the limitations of existing models by integrating a dynamic lexicon-based approach with advanced machine learning techniques. This hybrid framework is designed to enhance both the accuracy and adaptability of threat detection while providing interpretability and scalability.

Dynamic Lexicon Integration: At the core of the proposed system is a dynamic lexicon that evolves in response to emerging political discourse. Unlike static lexicons, this dynamic approach utilizes unsupervised learning techniques, such as topic modeling and clustering, to continuously update and expand the lexicon based on real-time data from social media, news sources, and other relevant platforms. By incorporating new terms and phrases, the system remains responsive to shifts in political language and emerging threats, ensuring that the lexicon accurately reflects current discourse and improves the system's detection capabilities.

Enhanced Machine Learning Models: The system leverages a suite of machine learning algorithms to analyze and predict political threats. It employs both traditional models, such as support vector machines (SVM) and decision trees, and advanced techniques like deep learning neural networks.

Context-Specific Adaptation: To address scalability and context-specific challenges, the proposed system incorporates region-specific and context-aware components. It uses localized lexicons and machine learning models tailored to different geographical areas and political environments. This adaptation ensures that the system accurately captures regional linguistic and cultural nuances, improving prediction accuracy for various contexts. Additionally, the system includes a modular architecture that allows for easy integration of new models and lexicons as needed, facilitating scalability and flexibility.

Improved Interpretability: Recognizing the importance of transparency, the proposed system integrates explainability features to address the interpretability challenges of machine learning models**Continuous Monitoring and Feedback:** The proposed system includes a feedback loop that enables continuous improvement and refinement. As new data is collected and analyzed, the system learns from its predictions and adjusts its models and lexicon accordingly.

4.1 Proposed system advantages:

1. Enhanced Adaptability: One of the primary advantages of the proposed system is its enhanced adaptability. By employing a dynamic lexicon that evolves based on real-time data, the system remains responsive to changes in political discourse and emerging threats

2. Improved Prediction Accuracy: The integration of both traditional machine learning models and advanced deep learning techniques contributes to significantly improved prediction accuracy. Traditional models, such as support vector machines (SVM) and decision trees, offer robustness and interpretability, while deep learning algorithms capture complex patterns and relationships within the data.

3. Scalability and Flexibility: The proposed system's modular architecture and context-specific adaptation features provide scalability and flexibility. The system is designed to handle multiple geographical areas and political environments by incorporating localized lexicons and tailored machine learning models. This ensures that the system can accurately address regional linguistic and cultural nuances, making it suitable for a wide range of contexts.

4. Enhanced Interpretability: Another significant advantage of the proposed system is its focus on interpretability. By combining lexicon-based sentiment analysis with machine learning predictions, the system provides clear explanations for its threat assessments

5. Continuous Improvement Through Feedback: The incorporation of a feedback loop in the proposed system ensures continuous improvement and refinement. As the system collects new data and receives user feedback, it learns from its predictions and adjusts its models and lexicon accordingly. This iterative process helps to address any discrepancies or gaps in threat detection, allowing the system to evolve and enhance its performance over time. By integrating real-world feedback and adapting to emerging challenges, the system remains effective and relevant, providing ongoing value for political security threat management.

V. SYSTEM IMPLEMENTATION

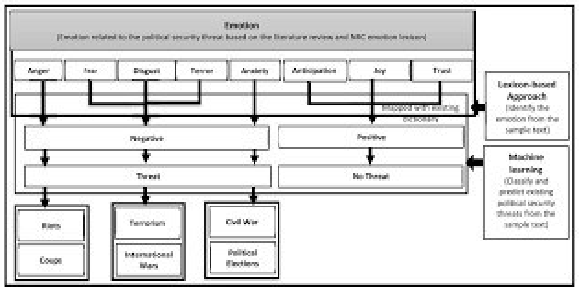
Data Processing and Integration: Data processing is a crucial aspect of the system. The framework begins by collecting and preprocessing data from diverse sources, including social media platforms, news feeds, and public statements. The dynamic lexicon filters this data, identifying relevant text segments based on political terms and sentiments. This preprocessing step ensures that the machine learning models receive high-quality, contextually relevant data.

Machine Learning Models: The system employs a combination of machine learning models to analyze the preprocessed data. Traditional models such as decision trees and random forests are used for their interpretability and robustness, while advanced techniques like deep learning neural networks capture complex patterns and relationships in the data.

Lexicon Management: A key feature of the proposed system is its dynamic lexicon management. Unlike static lexicons, the dynamic lexicon is continually updated using unsupervised learning techniques such as topic modeling and clustering. This approach allows the lexicon to adapt to emerging political language and trends, ensuring that it remains relevant and effective.

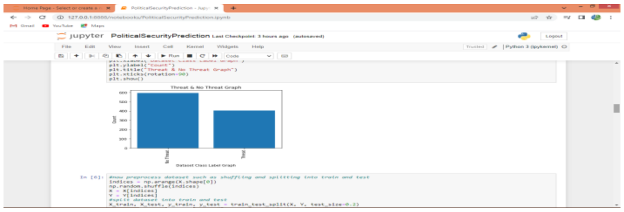
Evaluation and Feedback: The framework includes mechanisms for ongoing evaluation and feedback, which are essential for maintaining and improving system performance. The system collects feedback from users and monitors its predictive accuracy over time.

VI. System Architecture:


Emotion
(Emotion related to the political security threat based on the literature review and NRC emotion lexicon)
Anger
Fear
Disgust
Terror
Anxiety
Anticipation
Joy
Trust
Mapped with existing dictionary
Lexicon-based Approach
(Identify the emotion from the sample text)
Negative
Positive
Machine learning
(Classify and predict existing political security threats from the sample text)
Threat
No Threat
Riots
Coups
Terrorism
International Wars
Civil War
Political Elections

VII RESULTS ANALYSIS

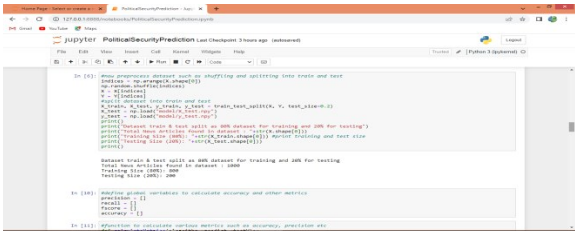
We have implemented this project using JUPYTER notebook and below are the code and output screens with blue colour comments



```
plt.ylabel("Count")
plt.title("Threat & No Threat Graph")
plt.xticks(rotation=90)
plt.show()
```

```
In [6]: #now preprocess dataset such as shuffling and splitting into train and test
        indices = np.arange(X.shape[0])
        np.random.shuffle(indices)
        X = X[indices]
        Y = Y[indices]
        #split dataset into train and test
        X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
```

In above screen finding and plotting graph of ‘Threat and NO Threat’ news found in dataset and in graph x-axis represents type of news and y-axis represents count



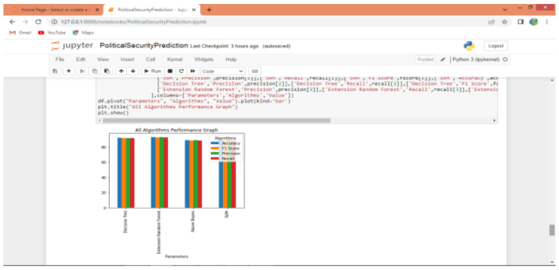
```
In [6]: #now preprocess dataset such as shuffling and splitting into train and test
        indices = np.arange(X.shape[0])
        np.random.shuffle(indices)
        X = X[indices]
        Y = Y[indices]
        #split dataset into train and test
        X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
        X_test = np.load("model/X_test.npy")
        y_test = np.load("model/y_test.npy")
        print()
        print("Dataset train & test split as 80% dataset for training and 20% for testing")
        print("Total News Articles found in dataset : "+str(X.shape[0]))
        print("Training Size (80%): "+str(X_train.shape[0])) #print training and test size
        print("Testing Size (20%): "+str(X_test.shape[0]))
        print()

        Dataset train & test split as 80% dataset for training and 20% for testing
        Total News Articles found in dataset : 1000
        Training Size (80%): 800
        Testing Size (20%): 200

In [10]: #define global variables to calculate accuracy and other metrics
         precision = []
         recall = []
         fscore = []
         accuracy = []

In [11]: #function to calculate various metrics such as accuracy, precision etc
```

In above screen processing and shuffling and splitting dataset into train and test



```
                ['SVM','Precision',precision[1]],['SVM','Recall',recall[1]],['SVM','F1 Score',fscore[1]],['SVM','Accuracy',acc
                ['Decision Tree','Precision',precision[2]],['Decision Tree','Recall',recall[2]],['Decision Tree','F1 Score',fs
                ['Extension Random Forest','Precision',precision[3]],['Extension Random Forest','Recall',recall[3]],['Extensio
               ],columns=['Parameters','Algorithms','Value'])
df.pivot("Parameters", "Algorithms", "Value").plot(kind='bar')
plt.title("All Algorithms Performance Graph")
plt.show()
```

VII. CONCLUSION

In conclusion, the proposed framework for predicting political security threats, which combines a hybrid lexicon-based approach with machine learning techniques, offers a comprehensive and nuanced method for analyzing and anticipating potential risks. The hybrid approach leverages the strengths of both lexicon-based sentiment analysis and advanced machine learning models, allowing for a more accurate and context-aware assessment of political discourse. By integrating established lexicons with machine learning algorithms, the framework can effectively capture the subtleties of political language and its implications for security.

The lexicon-based component provides a robust foundation by categorizing and interpreting political sentiments and key phrases that are often indicative of security threats. This is complemented by machine learning techniques that enhance the framework’s ability to adapt and learn from new data, improving its predictive capabilities over time. The synergy between these methods ensures that the system remains relevant and responsive to evolving political contexts and emerging threats.

VIII FUTURE WORK

For future work, several avenues can be explored to enhance the Political Security Threat Prediction Framework that utilizes a hybrid lexicon-based approach and machine learning techniques. First, expanding the lexicon to include a broader range of political terms, jargon, and emerging slang could improve the framework's sensitivity and accuracy. This expansion would involve continuously updating the lexicon to reflect current political discourse and regional variations, ensuring that the system remains relevant across different contexts and geographies.

Second, incorporating advanced machine learning techniques such as deep learning models and ensemble methods could further enhance the predictive power of the framework. These techniques have shown promise in capturing complex patterns and nuances in textual data, which could improve the system's ability to identify subtle or emerging threats. Experimenting with different architectures, such as transformers and recurrent neural networks, may yield more sophisticated insights and better performance.

References:

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(Jan), 993-1022. This seminal paper introduces Latent Dirichlet Allocation (LDA), a generative model that has been widely used for topic modeling in text analysis. The insights from this work can be applied to understanding political discourse and predicting security threats based on topic distributions.

Cheng, J., & Bernstein, M. S. (2015). Measuring and Modeling the Impact of Political Polarization on Online Communities. Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems. This study explores the effects of political polarization in online communities, providing a foundation for understanding how sentiment and discourse in such platforms can be indicative of potential security threats.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers. Liu's comprehensive work on sentiment analysis and opinion mining offers crucial methodologies and techniques that are integral to the lexicon-based component of threat prediction frameworks.

Manning, C. D., &Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press. This foundational text on statistical natural language processing provides essential background knowledge for implementing machine learning techniques in text analysis, which is relevant to the hybrid approach used in the framework.

Author's Profile:

S.S.Rajakumari, She Have Total 10 Years Of Teaching Experience Currently She Is Working As Associate Professor In St.John's College 0f Engineering And Technology. Completed M.Tech Under Jntu Anantapur, A.P, India

G.Sudha, completed B.Tech in CSE in St.John's College 0f Engineering And Technology ,Yerrakota in 2023,pursuing M.Tech in CSE in St.John's College 0f Engineering And Technology ,Yerrakota , Affiliated To Jntu Anantapur, India