

Sentinel: An Integrated Model for Predicting Political Security Threats

M.Anitha¹, K.Hareesh²,B.Vasavi Swarajya Lakshmi³

#1 Assistant & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor in the Department of MCA,SRK Institute of Technology, Vijayawada

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

ABSTRACT_ This project focuses on monitoring online sentiments and opinions to enhance national security. Excessive emotions expressed online can potentially lead to threats like riots and civil unrest, which jeopardize social and political stability. Researchers highlight the connection between emotions, sentiments, and political security risks. To address this, the project introduces a novel framework that predicts political security threats using a hybrid approach combining lexicon-based analysis and machine learning in cyberspace. Decision Tree, Naive Bayes, Support Vector Machine classifiers and Random Forest classifier are employed. Results demonstrate that the hybrid Lexicon-based approach with Random Forest classifier achieves the highest performance in predicting political security threats, emphasizing the framework's effectiveness.

1.INTRODUCTION

Cyberspace has emerged as a critical domain in the realm of national security, with its significance highlighted in various intelligence reports. The Worldwide Threat Assessment of the US Intelligence Community (2016) underscores cyber-related threats as being on par with other major concerns such as terrorism, weapons proliferation, and counterintelligence. Safeguarding a nation's security in the modern era has grown increasingly complex, thanks to the deluge of big data, widespread dissemination of information, and the proliferation of online rumors and fake news.

Interconnected nature of cyberspace means that negative emotions and disruptive behaviors can quickly spread, posing significant risks to national security. Research has shown a clear correlation between emotional triggers and the

emergence of security threats. Negative sentiments expressed in online content have the potential to evoke feelings of anger or fear, which in turn can escalate into events detrimental to national security. Given the prevalence of emotionally charged content in cyberspace, real-time detection of disruptive emotions is crucial for early intervention and effective management by authorities.

Despite the evident importance of understanding and managing emotions, there remains a notable space in research regarding the assessment and measurement of emotions. While opinion mining techniques have been explored in other domains, their application to national security, particularly in the context of emotion detection, remains underdeveloped. Consequently, there is a pressing need for comprehensive research in this area to better identify and address emerging threats.

Existing studies have primarily focused on classifying human emotions using various methodologies, overlooking the nuanced relationship between emotions and national security threats. In response, this study proposes a clever hypothetical system for foreseeing political dangers in light of the examination of feelings implanted inside web-based news content. Political security, as a urgent part of public safety, fills in as the point of convergence of this exploration.

Proposed work is authenticated through experimental study utilizing a hybrid approach that combines lexicon-based-methods with ML techniques, including DT, Naïve Bayes, and SVM. By integrating sentiment analysis with emotion detection, the framework aims to provide a more nuanced understanding of political threats and their emotional underpinnings.

Experimental process involves gathering text data from online news platforms and subjecting it to various machine learning algorithms to assess performance, accuracy, and precision. The results of the experiments offer insights into the efficacy of different hybrid methods in predicting political threats based on emotional cues present in online news content

3.LITERATURE SURVEY

A. Balahur *et al*

This article conducts a comparative analysis of techniques for opinion mining from quoted speech in newspaper articles. It highlights the complexity of the task due to diverse targets and affect phenomena within quotes. Evaluation with annotated data from the EMM news gathering engine

reveals the necessity of large lexicons and specialized training data for effective opinion mining systems.

N. Razali *et al*

Opinion mining, a subset of (NLP), delves into extracting sentiments, attitudes, and emotions from text. In the digital age, where public opinion drives decisions, it's a vital yet complex task. Recent research has seen advancements in sentiment analysis techniques, from machine learning to lexicon-based approaches. However, gaps still exist, paving the way for further exploration and innovation in sentiment analysis methodologies and their application across various domains in cyberspace.

S. Taj *et al*

Rapid evolution of Information Technology has revolutionized news dissemination, inundating platforms with vast amounts of user-generated data. Traversing this data manually is daunting, necessitating automated methods like sentiment analysis. This study introduces a lexicon-based approach for analyzing sentiments in news articles. Experiments conducted on the BBC news dataset validate the efficacy of this method, showcasing its relevance in navigating the emotional landscape of modern media.

M. Yassine *et al*

This paper proposes a novel framework to analyze emotional interactions in online social networks, aiming to differentiate between friends and acquaintances based on emotional content. Text mining techniques are employed to extract emotions from comments on social platforms, considering the informal

language used. The framework includes data collection, processing, and mining steps, along with specialized lexicons for online language. Through unsupervised techniques like k-means clustering, the model demonstrates high accuracy in discerning text subjectivity and predicting friendship dynamics, with Lebanese Facebook users as a case study.

T. G. Coan *et al*

This research delves into the emotional and coping responses of employees facing cyber-attacks, crucial in today's IT-dependent organizations. Through a case study at a global manufacturing firm, data from 24 interviews, observations, and secondary sources inform grounded analysis. Applying Technology Threat Avoidance Theory (TTAT), findings reveal IT security teams employing both positive problem-solving and negative emotion-focused coping strategies. Senior management's empathetic support emerges as a pivotal factor, shaping coping dynamics. The study introduces the Transformation of Coping through Empathic Leadership (T-CEL) model, highlighting management's role in bolstering employee resilience.

3. PROPOSED SYSTEM

To detect such news author of this paper 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

learning algorithm to train a model and this can be applied on any news to predict Threat or No Threat Label.

In propose work author has used SVM, Naive Bayes and Decision Tree as the machine learning algorithm and each algorithm performance is evaluated in terms of accuracy, precision, recall and F-SCORE. Decision Tree is giving best accuracy among all. And as extension we have experiment with Random Forest classifier is giving best accuracy compare to all exiting methods

3.1 IMPLEMENTATION

3.1.1 Data Collection and Preprocessing

Import necessary Python libraries including NLTK, scikit-learn, pandas, and numpy.

Load the news dataset from a CSV file named "News.csv".

Dataset contains columns such as news_index, id, text, date_created, and label.

3.1.2 Preprocess the news data by cleaning the text:

Remove special symbols and punctuation.

Tokenize the text.

Remove stopwords and non-alphabetic characters.

Stem and lemmatize the words.

The cleanNews() function preprocesses the text data by removing special symbols, punctuation, and stop words while also performing stemming and lemmatization.

3.1.3 NRC Lexicon emotion analysis

calculateNRCLexicon() function uses the NRC Lexicon to assign emotion labels ("Threat" or "No Threat") to each news article in the dataset.

NRC Lexicon emotion analysis is a vital role in shaping the sentiment of news articles, which is then used to train and evaluate ML models. Models' performance metrics indicate their effectiveness in classifying news articles based on threat levels. Label the news dataset using the NRC Lexicon emotion analysis to determine whether each news article represents a "Threat" or "No Threat" on the positive, negative scores.

3.1.4 Feature Engineering

(TF-IDF) vectorizer to convert text data into numerical feature vectors

Save the TF-IDF vectorizer for future use.

3.1.5 Data Splitting

Dataset is split into training and testing sets for model evaluation. Split ratio is 80% for training and 20% for testing. This step ensures that the model's performance can be evaluated.

3.1.6 Model Training and Evaluation

Naive Bayes Classifier:

Train a Naive Bayes classifier on the TF-IDF vectors.

Evaluate the metrics such as accuracy, precision, recall, and F1-score.

Save the trained Naive Bayes classifier for future use.

(SVM) Classifier:

Train an SVM classifier on TF-IDF vectors.

Estimate the classifier using performance metrics.

Save the trained SVM classifier for future use.

Decision Tree Classifier:

Train a Decision Tree classifier on TF-IDF vectors.

Estimate the classifier using performance metrics.

Save the trained Decision Tree classifier for future use.

Random Forest Classifier:

Train a Random Forest classifier on the TF-IDF vectors.

Evaluate the classifier using performance metrics.

Save the trained Random Forest classifier for future use.

3.1.7 Model Comparison and Analysis

Analyze the precision, recall, F1-score, and accuracy of each classifier.

Compare the performance of all the trained classifiers using a bar graph.

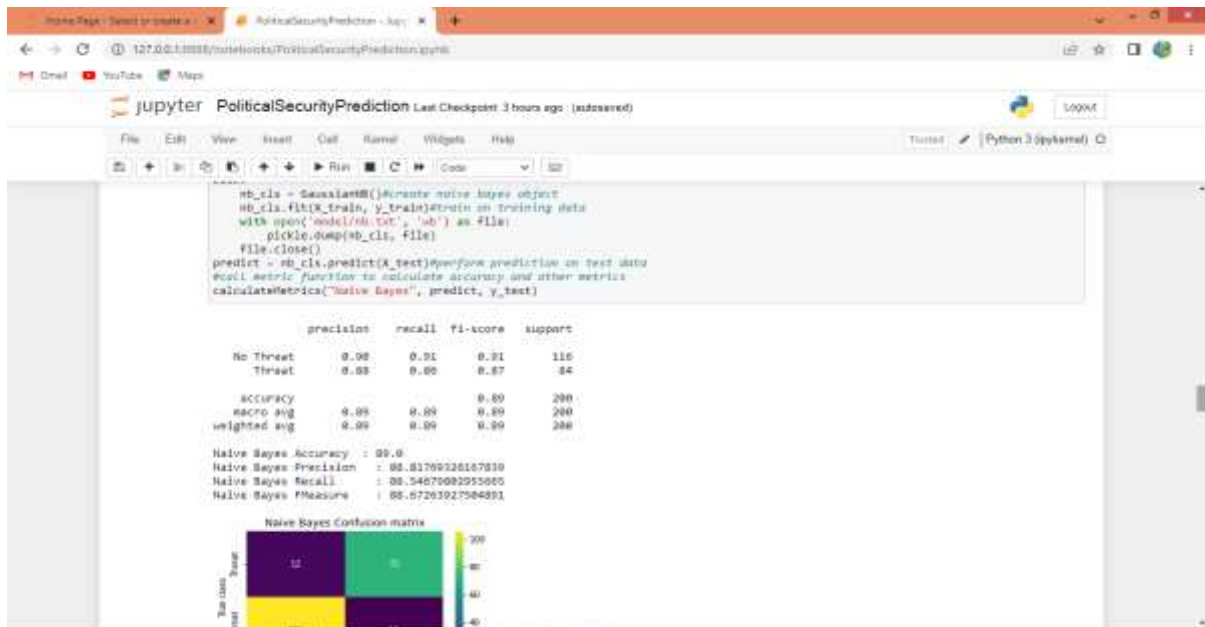
3.1.8 Test Data Prediction

Load test data from a CSV file named "testData.csv".

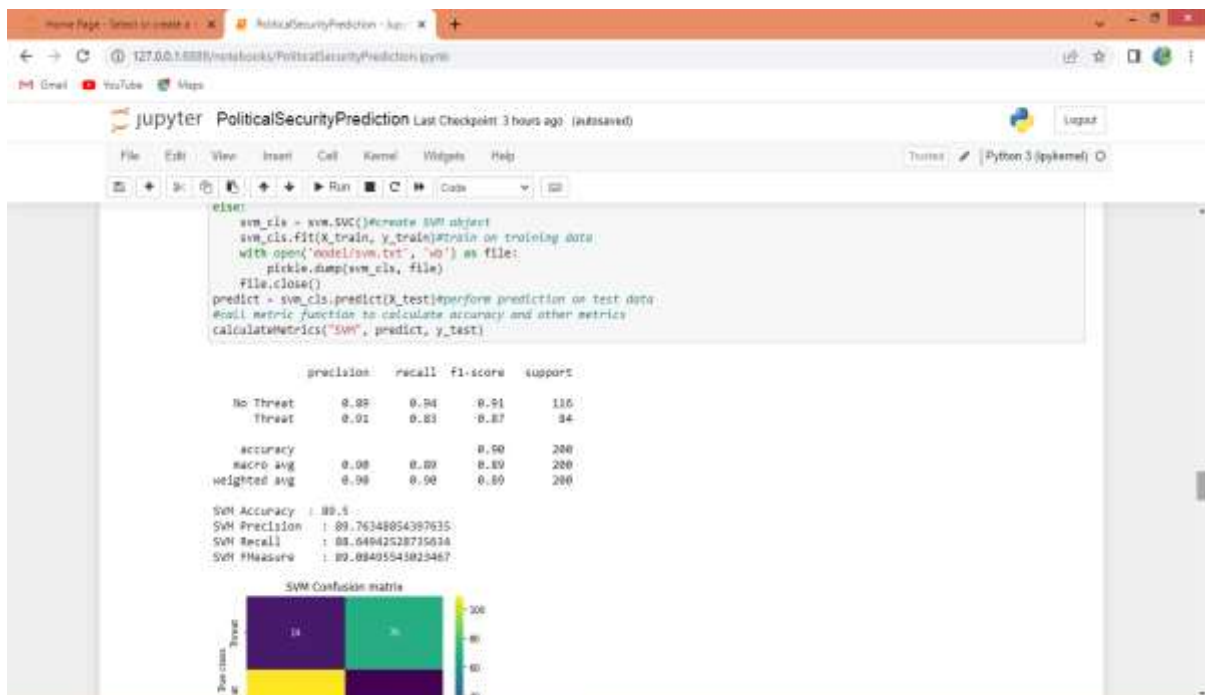
Use trained Random Forest classifier to predict the threat level of the test news articles.

Display the predicted threat level for each test news article.

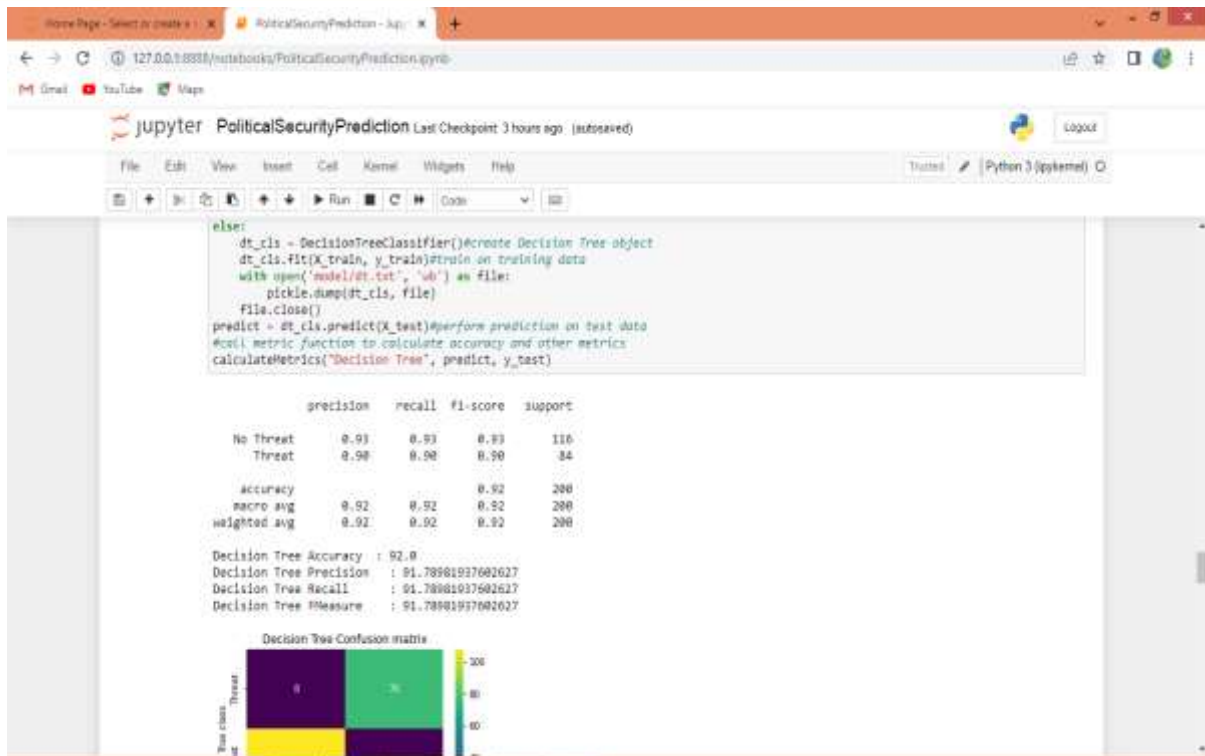
4.RESULTS AND DISCUSSION



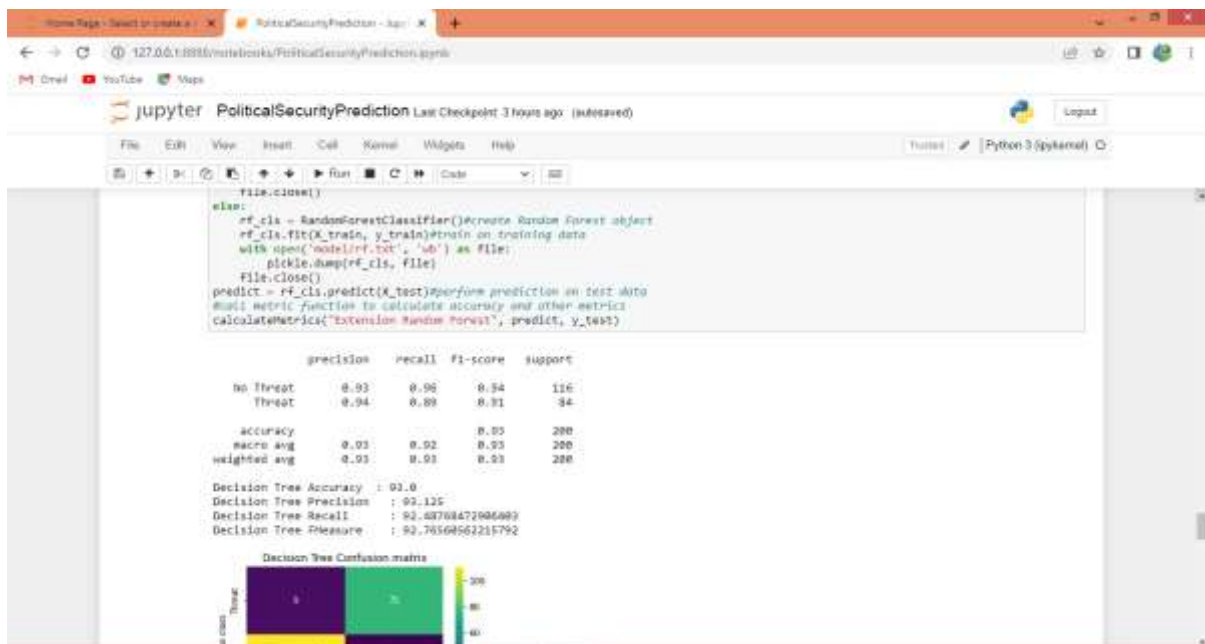
Training Naive Bayes and then displaying accuracy



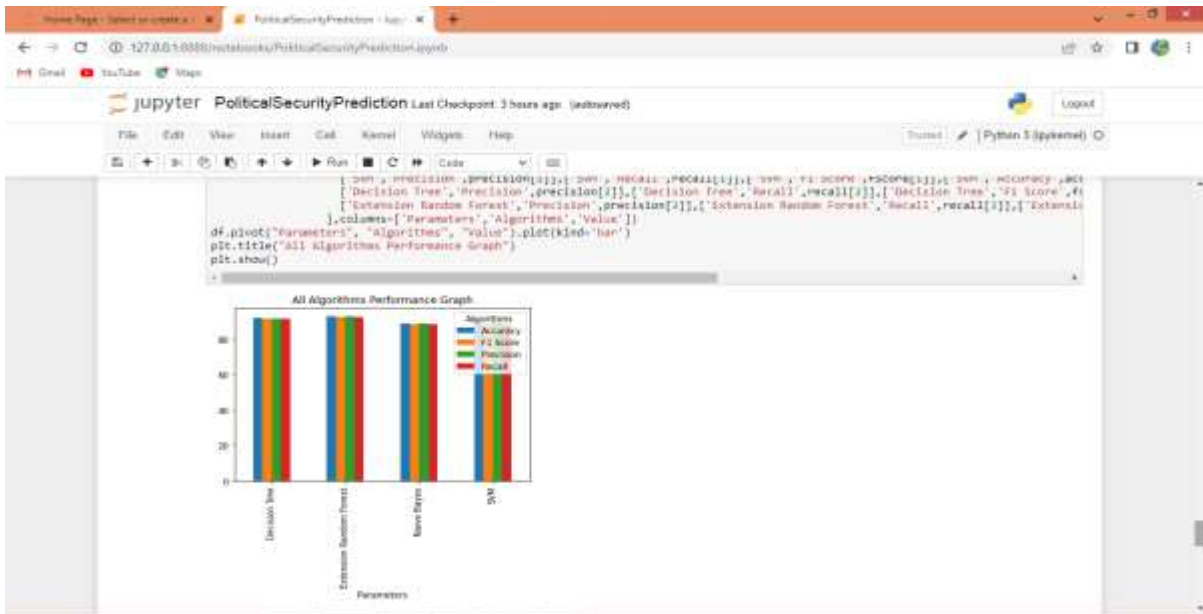
SVM got 99.50% accuracy and can see other output also



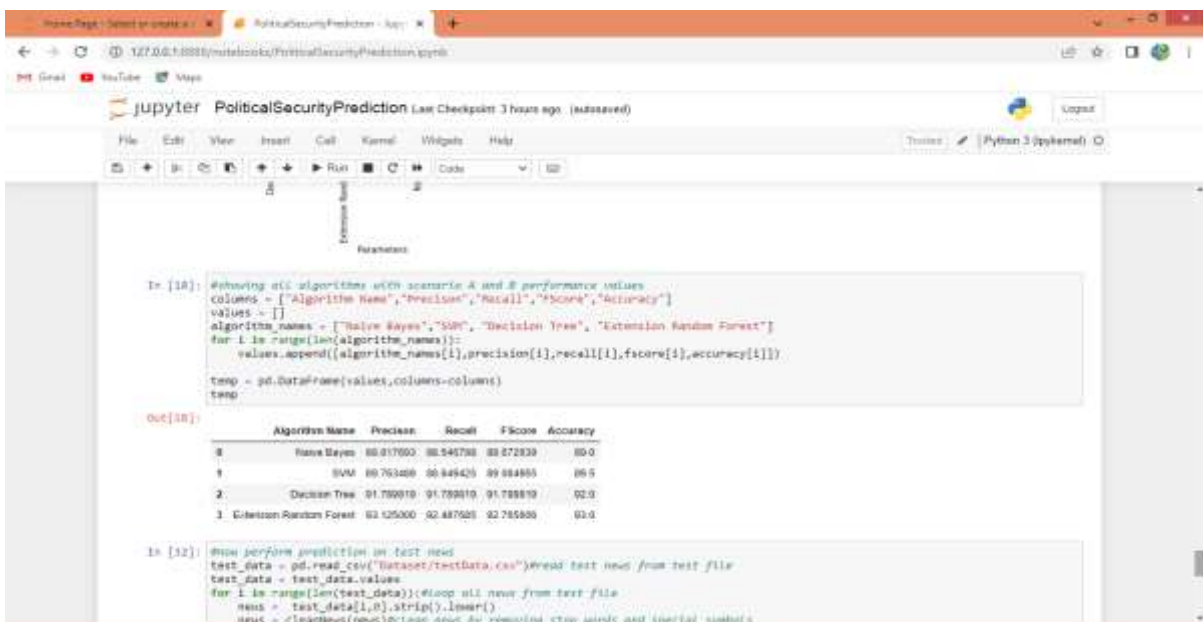
Decision tree got 92% overall accuracy



Extension Random Forest got 93% overall accuracy



Graphical representation for accuracy, other metrics of all algorithms



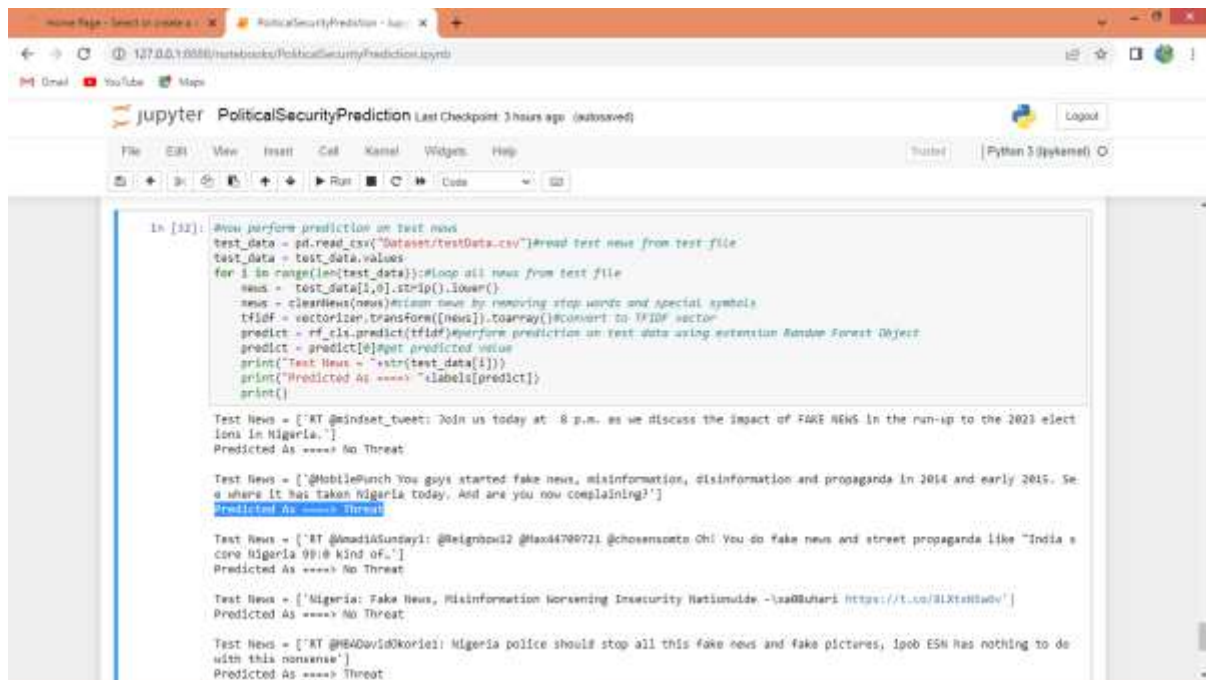
```
In [18]: #showing all algorithms with scenario A and B performance values
columns = ["Algorithm Name", "Precision", "Recall", "F1 Score", "Accuracy"]
values = []
algorithm_names = ["Naive Bayes", "SVM", "Decision Tree", "Extension Random Forest"]
for i in range(len(algorithm_names)):
    values.append([algorithm_names[i], precision[i], recall[i], f1score[i], accuracy[i]])

temp = pd.DataFrame(values, columns=columns)
temp
```

	Algorithm Name	Precision	Recall	F1 Score	Accuracy
0	Naive Bayes	00.817600	00.846700	00.829300	00.0
1	SVM	00.763400	00.849400	00.804900	00.5
2	Decision Tree	01.780910	01.780910	01.780910	02.0
3	Extension Random Forest	02.125000	02.125000	02.125000	02.0

```
In [12]: #now perform prediction on test news
test_data = pd.read_csv('dataset/testData.csv')#read test news from test file
test_data = test_data.values
for i in range(len(test_data)):#loop all news from test file
    news = test_data[i,0].strip().lower()
    news = clearNews(news)clear news by removing stop words and special numbers
```

Displaying all algorithm performance in tabular format



```

In [12]: #You perform prediction on test news
test_data = pd.read_csv('Datasets/testData.csv')#read test news from test file
test_data = test_data.values
for i in range(len(test_data)):#loop all news from test file
    news = test_data[i,0].strip().lower()
    news = cleanNews(news)#clean news by removing stop words and special symbols
    tfidf = vectorizer.transform([news]).toarray()#convert to TFIDF vector
    predict = rf_fit.predict(tfidf)#perform prediction on test data using extension Random Forest Object
    predict = predict[0]#get predicted value
    print('Test News = '+str(test_data[i]))
    print('Predicted As ==>'+ str(labels[predict]))
    print()

Test News = ['RT @indset_tweet: Join us today at 8 p.m. as we discuss the impact of FAKE NEWS in the run-up to the 2023 elect
ions in Nigeria. #']
Predicted As ==> No Threat

Test News = ['@MobilePunch You guys started fake news, misinformation, disinformation and propaganda in 2014 and early 2015. Se
e where it has taken Nigeria today. And are you now complaining?']
Predicted As ==> Threat

Test News = ['RT @madiaSunday1: @Reignopol2 @Max44709721 @chosensento Oh! You do fake news and street propaganda like "India s
core Nigeria 99% kind of..']
Predicted As ==> No Threat

Test News = ['Nigeria: Fake News, Misinformation worsening Insecurity Nationwide -Laa@Luhari https://t.co/3LXtst8sdv']
Predicted As ==> No Threat

Test News = ['RT @B4Davis@koriel: Nigeria police should stop all this fake news and fake pictures, Ijob ESN has nothing to do
with this nonsense']
Predicted As ==> Threat

```

Predicted output as ‘Threat or No Threat’

5. CONCLUSION

This project addresses the critical need to monitor and mitigate potential threats arising from public sentiments expressed on social media. By harnessing a hybrid approach combining lexicon-based analysis with machine learning algorithms, the framework effectively identifies and predicts the likelihood of political unrest or threats based on emotional cues extracted from public discourse. The use of advanced algorithms such as SVM, Naive Bayes, Decision Tree, and the extension to Random Forest demonstrates promising results in terms of accuracy and performance. Through proactive monitoring and intervention, this framework offers a proactive strategy for governments to prevent civil unrest, riots, and ultimately foster political stability and security.

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.

AUTHOR’S PROFILE



Ms.M.Anitha Working as Assistant & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with B .tech, MCA ,M. Tech in Computer Science .She has 14 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.



Mr.K.Hareesh completed her Master of Computer Applications. Currently working as an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. His area of interest includes Networks, Machine Learning.



Ms.B.Vasavi Swarajya Lakshmi is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc.(statistics) from P.R.Government college, kakinada, ,East Godavari District. Her area of interest are DBMS and Machine Learning with Python.