CYBER THREAT DETECTION AND PROFILING USING AI

[1] Anki Shah, [2] J Kathyayani, [3] D Janani, [4] E Abhinav, [5] Rajashree Sutrawe
[1], [2], [3], [4] UG Scholars, [5] Associate Professor, Department of computer science and Engineering,
Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India
*Corresponding author E-mail: kathyayanijala763@gmail.com

ABSTRACT
The increasing reliance on the internet has escalated the frequency and sophistication of cyber threats, making timely identification and mitigation essential. This research presents an AI-powered framework for cyber threat detection and profiling using Natural Language Processing (NLP) and Machine Learning (ML) techniques. By utilizing Twitter as an Open Source Intelligence (OSINT) platform, the system collects real-time threat intelligence, classifies threats, and maps them to the MITRE ATT&CK framework to provide actionable insights. Key processes include data preprocessing, feature extraction using advanced NLP models, and threat profiling to assess intent, origins, and potential impacts. The automated approach reduces analyst workload, enhances accuracy, and accelerates response times, addressing the limitations of manual threat analysis and noisy data sources. This framework aims to advance proactive cybersecurity by delivering real-time, context-aware threat intelligence.

Keywords: Cybersecurity, Threat Intelligence, Natural Language Processing (NLP), Machine Learning (ML), Open Source Intelligence (OSINT), Twitter, MITRE ATT&CK, Threat Detection, Threat Profiling, Real-Time Alerts, Cyber Threats, Automated Analysis, POS Tagging, Spacy

I. INTRODUCTION:

As the cyber threat landscape rapidly evolves, the shrinking window between vulnerability disclosure and exploitation poses a critical challenge. High- profile cases like the Log4j vulnerability highlight how quickly attackers can launch ransomware and cryptocurrency mining campaigns. This urgency calls for proactive cybersecurity strategies capable of detecting and understanding the threats in real time.

To address this, we propose a novel framework for automatic cyber threats identification and profiling by using Twitter as a real-time event source and the MITRE ATT&CK framework for threat characterization. The system consists of three core components: identifying threats and their names, profiling their intent through a layered machine learning model, and generating risk-based alerts.

Our approach enhances traditional detection by offering deeper insights into threat intentions, enabling timely and targeted mitigation. In experiments, the profiling component achieved 77% F1 score, demonstrating strong performance in characterizing emerging threats. This work contributes to proactive cyber defense by combining social media intelligence with advanced machine learning for early and informed threat response.

II. MODULES

A. SERVICE PROVIDER
The Service Provider logs in with valid credentials to perform various tasks, including training and testing user profile datasets, viewing accuracy results (including bar charts), analyzing identity prediction ratios, downloading predicted datasets, and managing remote users.

B. VIEW AND AUTHORIZE USERS
The admin can view the list of registered users along with their details (username, email, and address) and authorize them for system access.

C. REMOTE USER
Multiple users can register and log in with valid credentials. Once logged in, users can perform actions such as predicting profile identity status and viewing their personal profile data.

III. EXISTING SYSTEM

Cybersecurity is increasingly critical, and Security Operations Centers (SOCs) require timely threat intelligence to stay effective. Manually analyzing vast information sources is inefficient, making Open Source Intelligence (OSINT)—including social media like Twitter—a valuable tool for detecting emerging cyber threats.

Several studies have proposed frameworks using OSINT and machine learning for threat detection. These approaches vary in methods, such as keyword filtering, CNN-based tweet classification, and novelty

detection using CVE data. However, most focus only on identifying cyber-related events or classifying known threat types, lacking capabilities for naming threats or profiling their behavior and intentions.

DRAWBACK

- Lack of Threat Naming
- Limited to Predefined Categories
- Keyword-Based Filtering Issues
- No Threat Profiling
- Lack of Real-Time Adaptability

IV. PROPOSED SYSTEM

Automatically identify and profile emerging cyber threats using OSINT (e.g., Twitter) to generate timely alerts for cybersecurity professionals.

A. PROPOSED APPROACH

STEP-1: Continuously monitor Twitter posts from key sources.
STEP-2: Apply Natural Language Processing (NLP) and Machine Learning (ML) to identify likely threat-related to terms.
STEP-3: Use MITRE ATT&CK to infer threat with tactics.
STEP-4: Generate alerts with threat profiles and risk ratings.

ADVANTAGES

- Detects threats early via OSINT.
- Profiles threats based on cyber kill chain phases.
- Helps analysts understand and respond faster.

B. MACHINE LEARNING MODELS USED

1. **Decision Tree:** Simple to understand, builds recursively by splitting data based on features.
2. **Gradient Boosting:** Builds models sequentially to reduce error, better than random forest in many cases.
3. **Logistic Regression:** Best for binary classification; doesn't assume normality like discriminant analysis.
4. **Naive Bayes:** Fast, easy to implement, works well despite assuming feature independence.
5. **Random Forest:** Combines multiple decision trees to avoid overfitting; works well with minimal tuning.
6. **SVM:** Finds the optimal boundary between classes, works well with high-dimensional data.

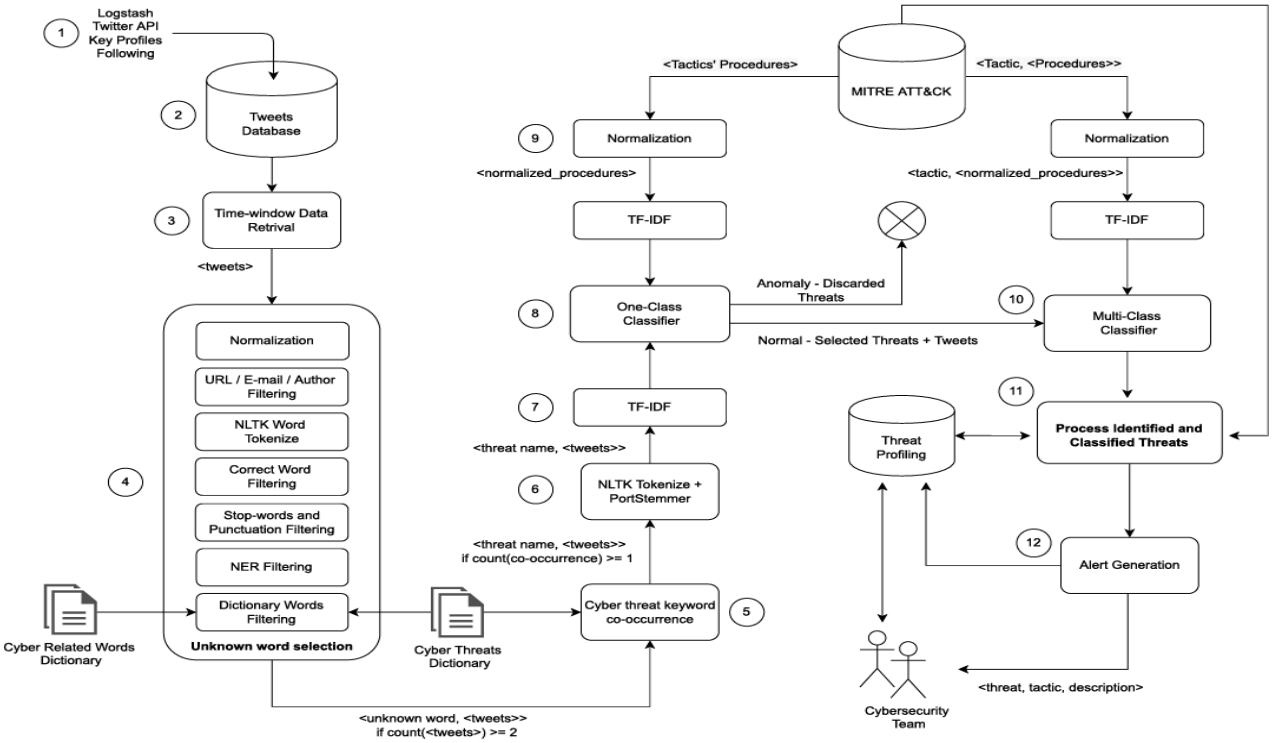

Fig.1.Proposed system

C. WORKING

1. **Data-Collection:**

Gather cyber threat-related text data from sources like Twitter, forums, news, and blogs.

2. **Preprocessing:**

Clean and process the text using NLP methods such as tokenization, stemming, lemmatization, and stop word removal.

3. **MITRE ATTACK Integration:** Map identified threats to known tactics and techniques using the MITRE ATT&CK framework.

4. **Model-Evaluation:**

Continuously validate and update ML models using new data to enhance prediction accuracy.

V. RESULT AND IMPLEMENTATION

Fig.2. Prediction of cyber threat type

Fig.3. Cyber threat found

VI. CONCLUSION:

This research presents an automated system for identifying and profiling emerging cyber threats using NLP on Twitter data. By mapping tweets to MITRE ATT&CK tactics, the system enhances situational awareness and supports early response. Real-world deployment demonstrated its effectiveness, with timely alerts enabling proactive defense.

VII. FUTURE ENHANCEMENTS:

Future work aims to improve the tweet selection stages to reduce false positives and enhance the accuracy of threat profiling. Experimentation with NLP techniques, such as the Part of Speech (POS) algorithm from the Spacy Python library, will help identify key phrases to refine the detection of unknown threats and improve overall performance.

REFERENCES

[1] B. D. Le, G. Wang, M. Nasim, and A. Babar, "Gathering cyber threat intelligence from Twitter using novelty classification," *arXiv preprint*, arXiv:1907.01755, 2019.

[2] Gartner Research, "Definition: Threat Intelligence," Stamford, CO, USA, 2013.

[3] R. D. Steele, "Open source intelligence: What is it? Why is it important to the military," *Journal*, vol. 17, no. 1, pp. 35–41, 1996.

[4] C. Sabottke, O. Suciu, and T. Dumitras, "Vulnerability disclosure in the age of social media: Exploiting Twitter for predicting real-world exploits," in *Proc. 24th USENIX Security Symposium (USENIX Security)*, 2015, pp. 1041–1056.

[5] A. Sapienza, A. Bessi, S. Damodaran, P. Shakarian, K. Lerman, and E. Ferrara, "Early warnings of cyber threats in online discussions," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 667–674.

[6] E. Nunes et al., "Darknet and deepnet mining for proactive cybersecurity threat intelligence," in *Proc. IEEE Conf. Intell. Secur. Informat. (ISI)*, Sep. 2016, pp. 7–12.

[7] S. Mittal, P. K. Das, V. Mulwad, A. Joshi, and T. Finin, "CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 860–867.

[8] A. Attarwala, S. Dimitrov, and A. Obeidi, "How efficient is Twitter: Predicting 2012 U.S. presidential elections using support vector machine via Twitter and comparing against Iowa electronic markets," in *Proc. Intell. Syst. Conf. (IntelliSys)*, Sep. 2017, pp. 646– 652.

[9] N. Dionísio, F. Alves, P. M. Ferreira, and A. Bessani, "Towards end-to-end cyber threat detection from Twitter using multi-task learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[10] O. Oh, M. Agrawal, and H. R. Rao, "Information control and terrorism: Tracking the Mumbai terrorist attack through Twitter," *Information Systems Frontiers*, vol. 13, no. 1, pp. 33–43, Mar. 2011.