e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

AUTOMATED EMERGING CYBER THREAT PROFILING AND ALERT SYSTEM

R. Sai Hanuman Koushik[*1], R. Akshay Kumar[*2], P. Chandradeep[*3], K. Gopichand[*4], Mrs. T. Suneetha[*5]

[*1,2,3,4]Student/Research Scholar, Department Of Cyber Security Malla Reddy University, Hyderabad Maisammaguda, Dulapally, Hyderabad, Telangana, India.

[*5]Associate Professor, Department Of Cyber Security Malla Reddy University, Hyderabad Maisammaguda, Dulapally, Hyderabad, Telangana, India.

ABSTRACT

Over time, new cyber security vulnerabilities emerge and cyber criminals exploit them in an increasingly smaller way. Recent events such as the Log4j vulnerability show positive aspects. A few hours after the vulnerability was reported, attackers began scanning the Internet for vulnerable hosts where they could distribute threats such as cryptocurrency miners and ransomware. Therefore, cybersecurity defense strategies must identify threats and their potential as quickly as possible to achieve protection success. While discovering new threats is important, it is difficult for security analysts because there is so much data and information that needs to be analyzed to find signs that a threat exists. In this sense, we propose a framework to identify and describe emerging threats, using Twitter messages as the source of events and Open Source Intelligence as a method aware of threat characteristics. The framework has three main components: Identifying cyber threats and their names; Analyzing threats by their target or targets, using two layers of machine learning to filter and classify tweets; and they can create threats based on their alarms. The main purpose of our work is a way to explain or describe threats to its purpose or objectives by providing more detail about the threats and their effects.

Keywords: Cyber Threat, Threat Detection, Alert, Threat Intelligence, Threat Profile.

I. INTRODUCTION

The rise of technology and the internet has brought about numerous benefits to individuals and organizations. However, with this advancement also comes the risk of cyber threats. Cyber threats have become increasingly sophisticated and prevalent, making it challenging for organizations to protect their systems and data. In response to this, there has been a growing need for an automated emerging cyber threat profiling and alert system. This paper will discuss the importance of such a system, its components, and how it can help organizations in mitigating cyber risks. However, alongside these advancements, there lurks an ominous shadow: the escalating risk of cyber attacks. These attacks, driven by various motivations ranging from financial gain to espionage, pose significant threats to individuals, organizations, and even nations. In response to this pressing challenge, the need for timely and accurate threat intelligence has become paramount. Cyber and intelligence can also be obtained from unofficial sources such as public blogs, the dark web, forums, and social media platforms. Illegal information allows a person or organization to post threatening information in natural language or inappropriate information on the internet. OSINT regarding cybersecurity is a source of early warning for cybersecurity events such as security breaches. To carry out a cyberattack, a malicious actor must 1) identify vulnerabilities, 2) acquire the tools and techniques needed to exploit those vulnerabilities, 3) select the target and find the actors, 4) create or purchase the necessary equipment, and 5) prepare. and join the fight. Other actors (administrators, security analysts, and even victims) may discuss the inadequacy of an attack or coordinate a response to the attack. These activities often take place online through social media, Internet forums (both open and dark), and professional blogs, and leave digital traces. Collectively, these digital signals provide a better understanding of changing cyber threats and can signal impending or ongoing attacks before the attack is detected at the target. For example, exploits are discussed on Twitter before being made public and even discussed on the darknet before being discussed on social media. Threat intelligence is evidence-based information that provides critical information about the context, mechanisms, countermeasures, impacts, and recommendations of existing or emerging threats in the cyber environment. In this context, our research addresses the growing problem of automated cyber threat analysis and warning

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

systems. Our mission is poised to transform the cyber threat intelligence landscape by leveraging the power of machine learning and leveraging the vast amounts of data provided by OSINT. Collectively, these digital signals provide a better understanding of changing cyber threats and can signal impending or ongoing attacks before the attack is detected at the target. For example, exploits are discussed on Twitter before being made public and even discussed on the darknet before being discussed on social media. Threat intelligence is evidence-based information that provides critical information about the context, mechanisms, countermeasures, impacts, and recommendations of existing or emerging threats in the cyber environment.

II. LITERATURE SURVEY

1. RENATO MARINHO, Since 2013, he has been teaching post-graduate malware analysis and incident response disciplines. His research interests include cyber threat visibility measurement, emerging cyber threat discovery, and malware analysis. (referred cyber threat detection and discovery)
2. RAIMIR HOLANDA, He has more than 100 publications in international journals, conferences, and book chapters. His main research interests include the IoT and sensor networks, blockchain, and cybersecurity. (referred to cyber security terms for word filtration.)
3. Niakanlahiji: Explored the role of machine learning in cyber threat intelligence, emphasizing the need for automated systems to handle the growing volume of data.
4. R. Campiolo: Investigated the efficacy of utilizing Twitter as an OSINT source for early threat detection, highlighting the potential of social media platforms in augmenting traditional threat intelligence sources.
5. T. Mitchell: Delved into the challenges posed by unstructured data in threat classification and proposed innovative techniques for effective analysis and interpretation.

III. METHODOLOGY

Our methodology encompasses a systematic and iterative approach to address the complexities inherent in automated emerging cyber threat profiling and alert systems. At the outset, data collection serves as the foundational step, wherein a diverse array of sources is tapped into to construct a comprehensive dataset reflective of the dynamic cyber threat landscape. Leveraging APIs and web scraping techniques, we harvest data from social media platforms such as Twitter, underground forums, blogs, and other OSINT repositories, ensuring a wide coverage of potential threat indicators. Following data collection, pre-processing assumes paramount importance in ensuring the quality and relevance of the dataset. Text normalization techniques are employed to standardize text data, while feature extraction methods facilitate the identification of key attributes indicative of cyber threats. Furthermore, data filtering mechanisms are implemented to eliminate noise and irrelevant information, thereby enhancing the efficacy of subsequent analysis and modeling efforts. The crux of our methodology lies in the development of robust machine-learning models capable of discerning patterns and trends within the collected data. Leveraging supervised learning paradigms, including Support Vector Machines (SVM), Random Forest, and deep learning architectures, we train our models on labeled datasets to recognize and classify emerging cyber threats. Moreover, our approach encompasses a holistic risk prediction analysis, wherein identified threats are subjected to rigorous scrutiny to assess their severity and potential impact. Through statistical methods, risk-scoring algorithms, and domain-specific rules, we evaluate the likelihood and consequences of various threat scenarios, thereby enabling stakeholders to prioritize response efforts and allocate resources judiciously.

- Data Collection: We gather data from diverse sources including Twitter, blogs, forums, and other OSINT platforms to construct a comprehensive dataset encompassing a wide array of cyber threat indicators.
- Preprocessing: The collected data undergoes rigorous preprocessing to ensure uniformity, consistency, and relevance. This involves text normalization, feature extraction, and filtering to remove noise and irrelevant information.
- Machine Learning Model Development: We employ state-of-the-art machine learning algorithms, such as Support Vector Machines (SVM) and Random Forest, to develop models capable of profiling emerging cyber threats based on their intentions, characteristics, and risk levels.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

- Risk Prediction Analysis: Finally, we conduct risk prediction analysis to assess the severity and potential impact of identified threats, thereby facilitating informed decision-making and proactive threat mitigation strategies.

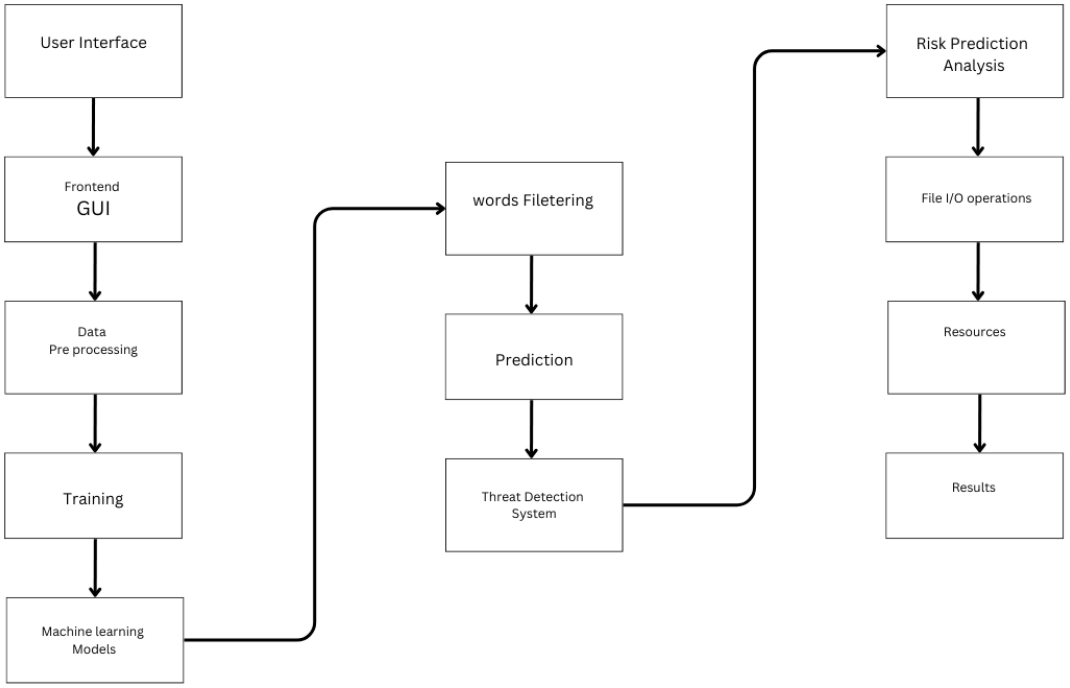

Figure 1: Architecture diagram

IV. SYSTEM ANALYSIS

Existing System

Cybersecurity has become a growing concern for many organizations, and there has been a lot of research in this area over the last few years. In these organizations, the Security Operations Center (SOC) is the central nervous system that provides the necessary security protection against cyber threats. This allows security analysts to effectively identify threats by collecting and reading multiple data sources. However, this is frustrating and cumbersome when done manually. OSINT is the collection, analysis, and use of information from publicly available information for intelligence purposes [21]. Examples of OSINT sites include public blogs, dark web sites, forums, and social media. On such a platform, any person or organization on the Internet can publish information regarding network security, including events, new threats, and vulnerabilities, in natural language. In the OSINT cyber threat intelligence site, we can say that Twitter is one of the most representative social media [20]. Cybersecurity experts, project managers, and hackers frequently use Twitter to discuss the topic of cyberattacks and share their experiences [4]. Different studies [1], [16], [8], [14], and [20] propose the use of OSINT to identify cyber threats through social media, forums, and other public spaces using text analysis. However, most of the recommendations focus on identifying critical situations related to cyber threats or vulnerabilities, but identifying and analyzing cyber threats is not recommended. Research [13] reported an early threat warning system that can scan online discussion content of online participants in social networks, security blogs, and the dark web to detect signs of potential cyber-attacks. The framework has two main components: paper mining and alert generation. The paper mining phase involves prioritizing input data to identify the threat list by discarding “known” content and equally selecting “unknown” content from different sources; because these may be new or discovered cyber threats. The second is stimulus generation, which is responsible for the stimulation of unknown words that meet certain requirements, for example, occurring twice. The method proposed in this study uses keyword filtering as the same strategy to identify cyber threats; This can lead to negative consequences as unknown words may appear in tweets or other non-cybersecurity-related content. Additionally, no cyber threats were detected in this research.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Proposed System

To solve the above problems, there should be an incentive for all customers to pay. This article describes an approach to a restaurant rating system that requires each customer to submit a rating after the visit to maximize ratings. It allows consumers to rate foods by eating them or by capturing images of their faces reflecting the corresponding emotion. The amount of data gathered from text-based evaluations is limited, and no personal data is obtained. However, a simple, fast, and fun rating system should provide an overall perspective on the customer's experience of the restaurant concept. Advantages of the application system:

- High efficiency
- Timesaving
- Continuous Monitoring
- Alert Generation

V. RESULTS

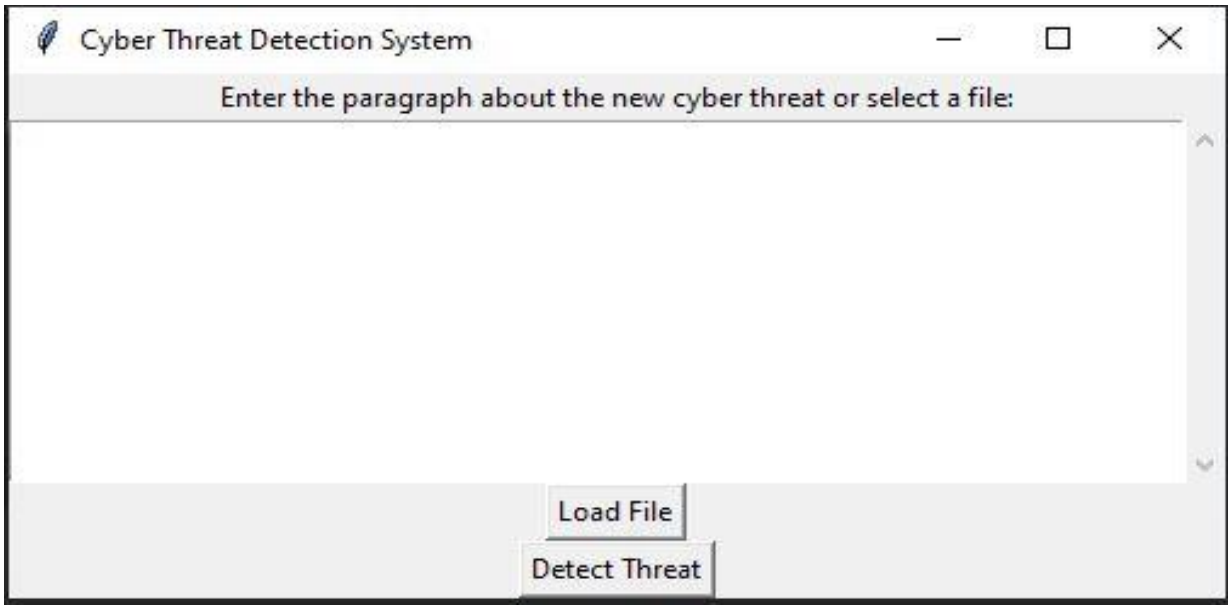

Figure 2: Cyber Threat Detection System

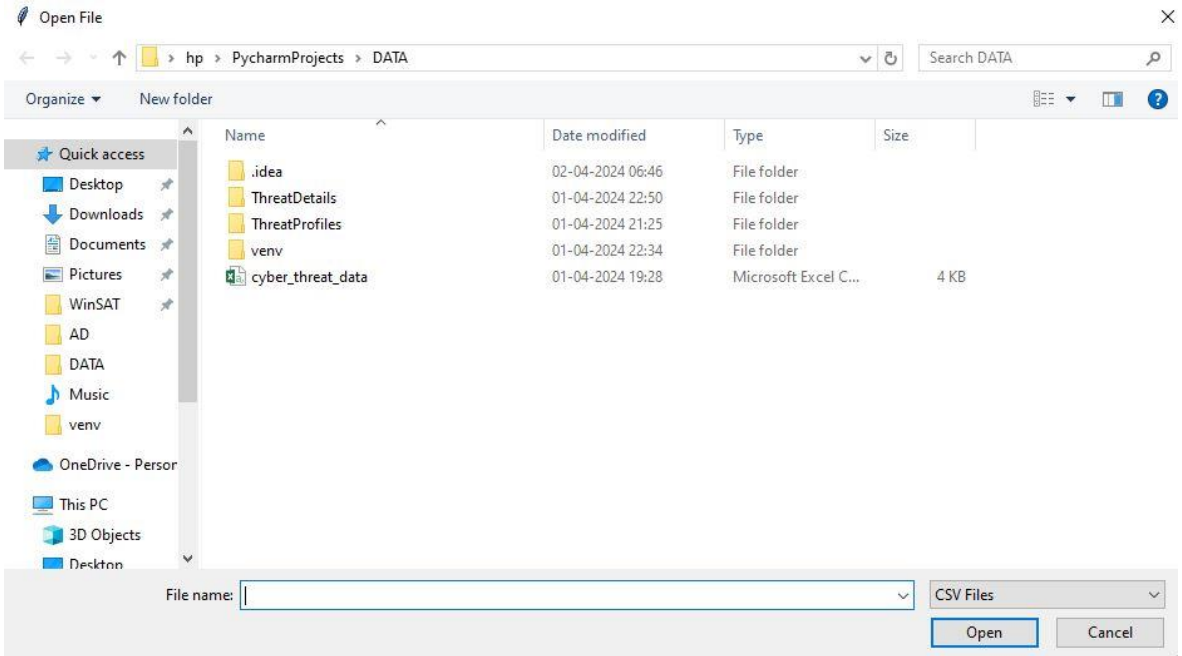

Figure 3: Load CSV file, text file, or paragraph as input

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

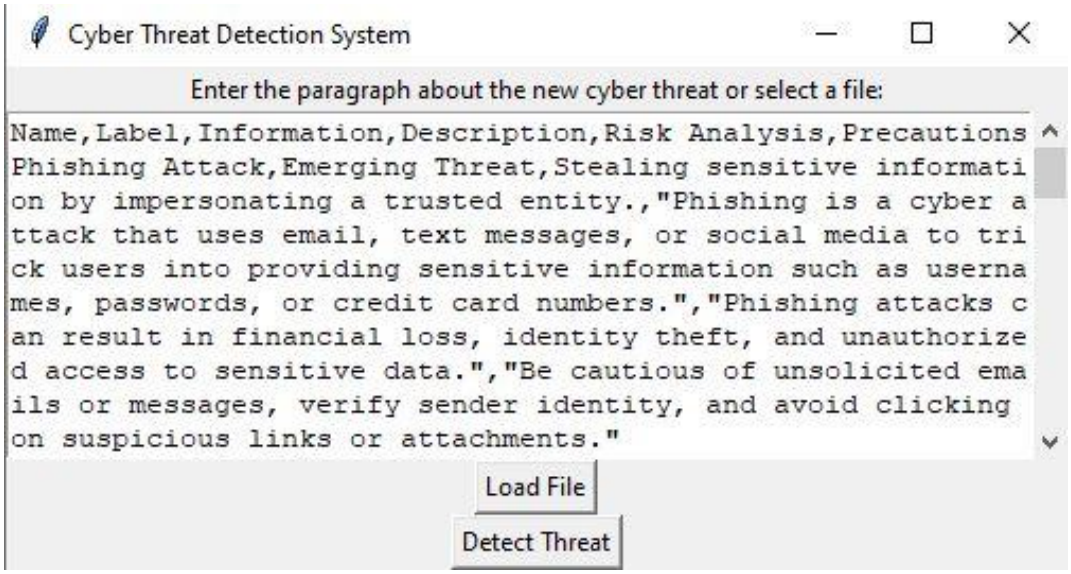

Figure 4: File has been successfully installed

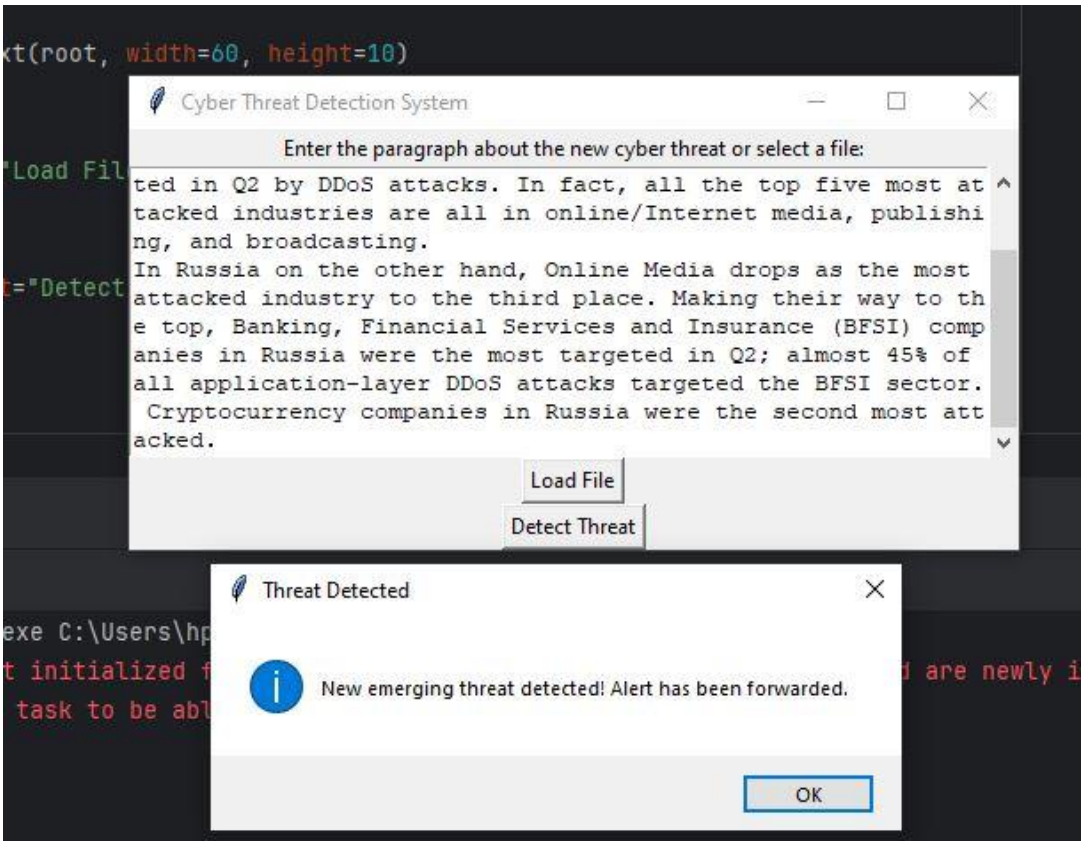

Figure 5: Threat detected from the input information

[**Figure 2**] Cyber Threat Detection System the Cyber Threat Detection System is a comprehensive framework designed to identify, analyze, and respond to cyber threats in real-time. It utilizes advanced technologies such as machine learning, data mining, and threat intelligence to detect and profile emerging cyber threats. The system continuously monitors various data sources, including social media platforms, forums, and OSINT repositories, for signs of potential threats. Upon detecting a threat, the system analyzes its characteristics, intentions, and risk level to assess its severity and potential impact. The Cyber Threat Detection System generates timely alerts and notifications to inform cybersecurity professionals and decision-makers about the detected threats, enabling them to take appropriate mitigation actions.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

[**Figure 3**] The "Load CSV File" functionality allows users to input data in CSV format into the Cyber Threat Detection System. Users can upload CSV files containing relevant information about cyber threats, including threat names, descriptions, and indicators. The system parses the CSV file and extracts the necessary data for further processing and analysis. This functionality facilitates the seamless integration of external threat data into the system, enhancing its threat detection capabilities.

[**Figure 4**] Upon uploading the CSV file or any other input data, the system processes the file and confirms successful installation. The system verifies the integrity of the uploaded file and ensures that all necessary data has been successfully extracted and imported. Users receive a confirmation message indicating that the file has been successfully installed and is ready for analysis. This step reassures users that their input data has been successfully integrated into the Cyber Threat Detection System and is available for further processing.

[**Figure 5**] After processing the input data, the Cyber Threat Detection System identifies and detects potential threats based on the provided information. Using advanced algorithms and threat intelligence, the system analyzes the input data to uncover patterns, anomalies, and indicators of cyber threats. Upon detecting a threat, the system generates an alert or notification to notify users about the identified threat. The alert includes details about the detected threat, such as its name, severity, and recommended actions for mitigation. Users can review the alert and take appropriate measures to respond to the detected threat, minimizing its impact on their organization's cybersecurity posture.

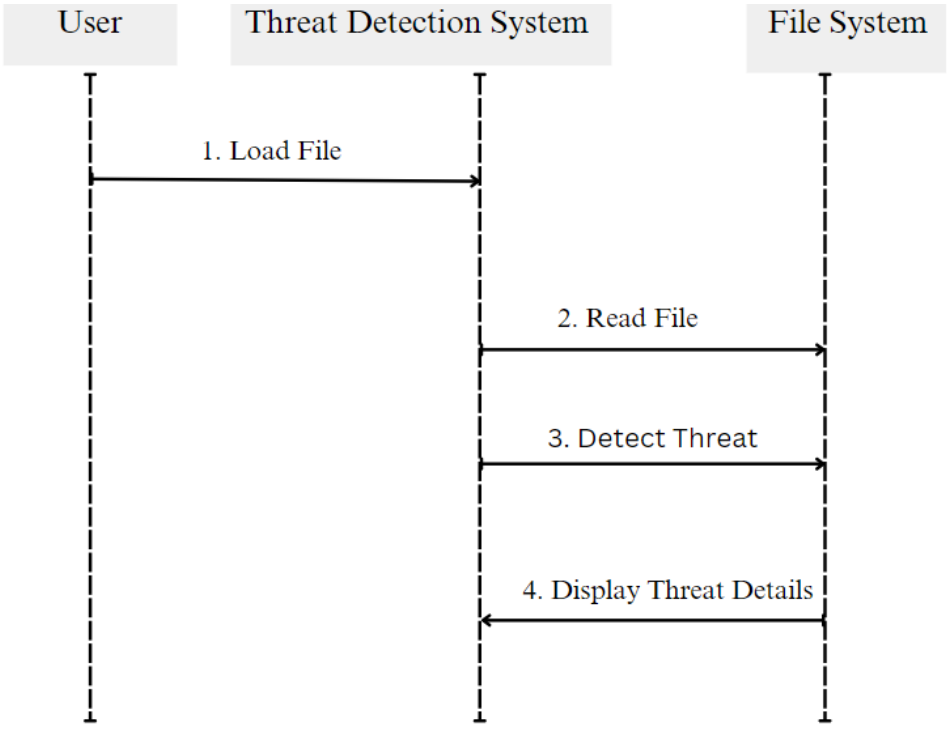

Figure 6: Sequence diagram of the system

VI. CONCLUSION

In conclusion, our research presents a comprehensive framework for automated emerging cyber threat profiling and alert systems, underpinned by advanced machine learning techniques and leveraging diverse sources of threat intelligence. Through systematic data collection, preprocessing, and modeling efforts, we have demonstrated the efficacy and potential of our approach in augmenting the capabilities of cybersecurity professionals and bolstering organizational resilience against evolving cyber threats. The results of our study underscore the transformative impact of automation and machine learning in enhancing cyber threat intelligence capabilities, enabling stakeholders to stay one step ahead of adversaries in the ever-changing threat landscape. By harnessing the power of data-driven insights and real-time analysis, our framework empowers organizations to proactively identify, assess, and mitigate emerging threats, thereby safeguarding critical assets and maintaining operational continuity in the face of cyber adversity.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

VII. FUTURE SCOPE

Looking ahead, our research paves the way for a multitude of future endeavors aimed at advancing the field of cyber threat intelligence and resilience. Key areas for further exploration and refinement include:

- Scalability and Efficiency: Efforts to enhance the scalability and efficiency of the proposed framework through optimization of data processing pipelines.
- Advanced Analytics Techniques: Exploration of advanced analytics techniques such as anomaly detection, predictive modeling, and behavioral analysis for proactive threat detection and response.
- Real-time Threat Intelligence: Integration of real-time threat intelligence feeds and automated threat hunting capabilities to enable continuous monitoring and adaptive response to emerging threats.
- Interpretability and Explainability: Research into methods for enhancing the interpretability and explainability of machine learning models, thereby facilitating trust, understanding, and adoption by cybersecurity practitioners.

VIII. REFERENCES

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

[2] Definition: Threat Intelligence, Gartner Research, 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 Secur. Symp. (USENIX Secur.), 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, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, J. Robertson, J. Shakarian, A. Thart, and P. Shakarian, ‘‘Darknet and Deepnet mining for proactive cybersecurity threat intelligence,’’ in Proc. IEEE Conf. Intell. Secur. Information. (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. VOLUME 11, 2023 58935 R. Marinho, R. Holanda: Automated Emerging Cyber Threat Identification and Profiling Based on NLP

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

[11] T. Sakaki, M. Okazaki, and Y. Matsuo, ‘‘Earthquake shakes Twitter users: Real-time event detection by social sensors,’’ in Proc. 19th Int. Conf. World Wide Web, Apr. 2010, pp. 851–860.

[12] B. De Longueville, R. S. Smith, and G. Luraschi, ‘‘‘OMG, from here, I can see the flames!’: A use case of mining location-based social networks to acquire spatio-temporal data on forest fires,’’ in Proc. Int. Workshop Location Based Social Netw., Nov. 2009, pp. 73–80.

[13] A. Sapienza, S. K. Ernala, A. Bessi, K. Lerman, and E. Ferrara, ‘‘DISCOVER: Mining online chatter for emerging cyber threats,’’ in Proc. Companion Web Conf. Web Conf. (WWW), 2018, pp. 983–990.

[14] R. P. Khandpur, T. Ji, S. Jan, G. Wang, C.-T. Lu, and N. Ramakrishnan, ‘‘Crowdsourcing Cybersecurity: Cyber attack detection using social media,’’ in Proc. ACM Conf. Inf. Knowl. Manage., Nov. 2017, pp. 1049–1057.

[15] Q. Le Sceller, E. B. Karbab, M. Debbabi, and F. Iqbal, ''SONAR: Automatic detection of cyber security events over the Twitter stream,'' in Proc. 12th Int. Conf. Availability, Rel. Secur., Aug. 2017, pp. 1–11.

[16] K.-C. Lee, C.-H. Hsieh, L.-J. Wei, C.-H. Mao, J.-H. Dai, and Y.-T. Kuang, ''Sec-buzzer: Cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation,'' Soft Comput., vol. 21, no. 11, pp. 2883–2896, Jun. 2017.

[17] A. Ritter, E. Wright, W. Casey, and T. Mitchell, ''Weakly supervised extraction of computer security events from Twitter,'' in Proc. 24th Int. Conf. World Wide Web, May 2015, pp. 896–905.

[18] A. Queiroz, B. Keegan, and F. Mtenzi, ''Predicting software vulnerability using security discussion in social media,'' in Proc. Eur. Conf. Cyber Warfare Secur., 2017, pp. 628–634.

[19] A. Bose, V. Behzadan, C. Aguirre, and W. H. Hsu, ''A novel approach for detection and ranking of trendy and emerging cyber threat events in Twitter streams,'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2019, pp. 871–878.

[20] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, ''Mitre ATT&CK: Design and philosophy,'' MITRE Corp., McLean, VA, USA, Tech. Rep. 19-01075-28, 2018.