International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

REAL TIME IDENTIFICATION OF EMERGING CYBER THREATS USING NLP-BASED MODELS

[#1]**Dr. RAMESH BOLLI,** ***Associate Professor, Department of CSE,***
[#2]**Dr. NALLA SRINIVAS,** ***Associate Professor, Department of CSE,***
[#3]**UPPULUTI RAGHU,** ***Department of CSE,***
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TG.

ABSTRACT: This research employs models based on Natural Language Processing (NLP) to analyze vast quantities of unstructured textual data from sources such as security reports, social media, dark web forums, and threat intelligence feeds in order to identify new cyber threats as they become apparent in real time. The suggested approach employs a variety of advanced natural language processing (NLP) methods, including semantic analysis, text classification, and named entity recognition, to identify patterns and early indicators of potential cyberattacks. In comparison to conventional signature-based methods, deep learning and machine learning models are capable of detecting threats at a faster and more precise pace by continuously learning from new data. The system enhances situational awareness by providing cybersecurity professionals with insightful data and instantly alerting them. In summary, the findings illustrate the effectiveness of NLP-driven frameworks in the identification and mitigation of emerging cyberthreats in real-time digital environments.

Keywords: *Real-Time Threat Detection, Cybersecurity, Natural Language Processing (NLP), Emerging Cyber Threats, Machine Learning, Deep Learning, Text Mining,*

1. INTRODUCTION

The capacity to identify new cyberthreats in real-time using NLP-based models is a critical component of modern cybersecurity, and it is becoming more critical as digital systems become more complex and interconnected. The exponential growth of internet use, online services, and cloud infrastructures has provided cybercriminals with a significantly larger target to exploit. Conventional security methods frequently overlook intricate and novel threats as a result of their substantial dependence on pre-established rules and signatures. The demand for intelligent systems that can detect threats in real time is high, as attackers are constantly evolving their strategies.

Natural Language Processing (NLP) can be implemented to effectively navigate the mountains of unstructured text data generated by the cybersecurity ecosystem. System logs, security reports, threat intelligence feeds, and internet forums can be employed to gain a more comprehensive understanding of potential cyberthreats. The categorization of harmful actions, pattern recognition, and feature extraction are all facilitated by natural language processing methods. Natural language processing models are advantageous for identifying concealed connections and cyberattack warning signs due to their capacity to convert unstructured text into structured insights.

The combination of natural language processing and machine learning

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

techniques enhances the accuracy and efficiency of threat detection systems. Transformer-based models, recurrent neural networks (RNNs), and long short-term memory (LSTM) networks are examples of advanced models that are capable of detecting subtle anomalies and comprehending contextual information. These models are highly effective at identifying evolving cyberthreats due to their perpetual learning from data.

In order to mitigate the effects of cyberattacks, we require real-time processing. Delays in detection can lead to a variety of severe repercussions, including financial losses, system disruptions, and data breaches. Real-time systems that are based on natural language processing utilize streaming data processing frameworks to monitor incoming data and send immediate alerts. By being proactive, responding promptly to events, and mitigating risks, organizations can prevent significant security breaches.

2. LITERATURE SURVEY

Rodriguez et al. (2021): This investigation introduces a real-time natural language processing (NLP) framework for the analysis of cybersecurity reports and data from the dark web. In order to extract pertinent threat information, the model employs topic modeling and named entity recognition. It devises novel methods to gain entry and exploit security vulnerabilities. By automating data analysis, the system enhances threat intelligence capabilities. The research emphasizes the importance of natural language processing in the development of preventative cyber defenses.

Ibrahim et al. (2021): Introduce a deep learning-based NLP model to identify cyberthreats in hacker communities and online forums. The writers employ recurrent neural networks to analyze textual conversations. The system detects suspicious patterns and new attack discussions. The accuracy of danger forecasting is improved by an understanding of the surrounding environment. The paper emphasizes the importance of deep learning in cybersecurity analytics.

Chatterjee et al. (2022): The objective of this investigation is to develop a fully automated cyber threat detection system that utilizes natural language processing and machine learning. The model analyzes vast quantities of text data related to security in real time. The objective of classification methods is to distinguish between data that is harmful and data that is harmless. One advantage of the system for threat analysis is the reduction of manual labor. The research illustrates the ways in which cybersecurity automation is improved by NLP.

Nguyen & Tran (2022): This investigation aims to extract cyber threat intelligence from unstructured data by employing transformer-based natural language processing models. The system employs sophisticated language models to contextualize threat data. It effortlessly identifies vulnerabilities and attack patterns. The model improves the accuracy and speed of detection. Modern NLP architectures are demonstrated to influence cyberthreat detection in this paper..

Almeida et al. (2023): For the purpose of identifying cyberthreats, the authors of this paper propose a hybrid natural language

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

processing framework that is based on machine learning and rules. The system evaluates security alerts and incident reports. Critical threat indicators that are extracted include the names of malware and attack techniques. The model's detection performance is improved by employing hybrid learning. The significance of employing a combination of NLP techniques to enhance results is underscored by the paper.

Das & Roy (2023): This paper delineates a system for real-time threat detection that employs analytics and natural language processing (NLP) on streaming data. The model is continuously analyzing data from cybersecurity feeds and logs. It employs techniques for the identification of entities and the detection of anomalies. The system responds promptly to the emergence of new threats. The paper indicates that real-time processing is a critical component of contemporary cybersecurity systems.

Fernandez et al. (2024): This paper introduces a model for the intelligent detection of cyber threats that employs NLP and knowledge graphs. The extracted entities are linked by the system through databases of recognized threats. This contextual information facilitates a more comprehensive comprehension of attack patterns and relationships. The model can enhance the quality of security analysts' decisions. This research demonstrates that semantic analysis can be a significant asset to threat intelligence.

Okafor & Bello (2024): This research employs natural language processing to develop a system for monitoring cyber threats in corporate environments. Internal logs, emails, and reports are examined by the system to detect potential hazards. Text clustering and classification methodologies are implemented for the purpose of analysis. The model enhances the organization's safety net. The primary focus of the research is the application of natural language processing to cybersecurity solutions for businesses.

Silva et al. (2025): This paper introduces a model for the real-time detection of new cyberthreats that is based on deep transformers. The system processes massive text datasets that originate from a variety of sources. It identifies attack patterns that have not been observed before and zero-day threats. Contextual embeddings enable the model to attain exceptional precision. The research concentrates on the evolution of cybersecurity transformer models.

Yadav & Mishra (2025): This investigation proposes a cyber threat detection system that integrates natural language processing with deep learning. The system employs a variety of models, such as CNN and LSTM, for text analysis. It detects malicious security-related patterns in text streams. The model improves detection performance and reduces false positives. The advantages of hybrid designs for security systems that employ natural language processing are examined in this paper.

Hassan & Qureshi (2026): This research develops an AI-powered cyber threat detection system using NLP and This research aimed to develop a cyber threat detection system that is AI-driven and utilizes reinforcement learning and natural language processing. The model adjusts in real time in response to changing patterns of danger. It continuously enhances the accuracy of detection through the

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

implementation of learning mechanisms. The system endorses proactive defensive strategies. The research underscores the evolution of intelligent cybersecurity systems over time.

3. PROPOSED ALGORITHM

In order to guarantee that cyber security engineers are promptly informed of any new cyberthreats, the primary objective of this research is to propose a method that can automatically identify and profile these threats through the use of OSINT (Open Source Intelligence). In order to achieve this objective, we provide a solution that integrates the subsequent macro steps.

1.Monitoring the Twitter activity of prominent accounts to detect any suspicious activity or cyberthreats by mining their posts for unknown terms;

2.The integration of machine learning (ML) and natural language processing (NLP) to identify potentially malicious terms and eliminate those that do not satisfy the criteria;

3.Utilize MITRE ATT&CK procedures to determine the most probable strategy for the identified threat.

4.Enhanced threat detection and response times by establishing risk-based objectives that are proportional to the rate of change in the threat since its discovery, threat characterization, and the development of early warning systems.

Key Components and Functionality:

Twitter Stream Collecto: Acts as the primary module responsible for data collection, retrieving tweets from cybersecurity sources in real time via the Twitter API.

Text Preprocessor: The text is prepared for analysis by cleaning, tokenizing, removing stop words, and normalizing it. ensures that the data is of high quality and consistent for processing.

Named Entity Recognizer (NER): Extracts IP addresses, attack types, malware names, vulnerability codes, and other entities related to threats. provides structured structures to disorganized writing.

Binary Classifier: Evaluates the cybersecurity relevance of tweets by employing supervised machine learning models, including SVM, CNN, and BERT. Content that is irrelevant is eliminated at the pipeline's inception.

Multi-Class Classifier: Associating relevant tweets with MITREATT&CK strategies and organizing them into distinct threat categories (such as ransomware and phishing). enables a thorough evaluation of potential hazards.

Risk Scoring Engine: The source's reliability, the frequency of the threats, and the novelty of the information are all taken into account when evaluating the gravity of the dangers discovered. To facilitate the prioritization of alerts and the formulation of mitigation decisions.

Database Management System: The system maintains a record of historical user data, threat profiles, raw data, and processed outputs. enables the execution of both immediate queries and longer-term analyses.

Visualization Dashboard: Displays analytics, trend charts, notifications, and threat maps in real time. enables users to interact with the results and navigate the data.

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

Alert Notification Module: Security analysts are informed of updates and alerts in accordance with predetermined thresholds. enables the immediate response to incidents.

User Access Control: Manages user privileges, authentication, and login. By granting access according to user roles, the system's functionality is safeguarded.

Advantages:

Automation of Threat Detection: The technology eliminates the necessity for human supervision of vast data sets by automatically evaluating tweets and identifying potential hazards. Consequently, productivity is enhanced as analysts are able to focus on more urgent matters.

Early Warning Capabilities: The system capitalizes on this fact by issuing early alerts, which facilitate faster response and mitigation, as social media platforms such as Twitter are frequently the first to identify cyber vulnerabilities and exploits.

Context-Aware Threat Profiling: The MITRE ATT&CK framework is employed by the system to profile threats by associating them with known tactics, techniques, and procedures (TTPs). Analysts can develop a more profound comprehension of the threat's behavior and the potential consequences by placing it in its appropriate context.

Reduction in False Positives: Traditional keyword-matching algorithms generate numerous false positives. Machine learning classifiers improve the precision of threat detection while simultaneously reducing the number of unnecessary alerts.

Scalability and Extensibility: The architecture is capable of accommodating businesses of any size due to its modular and scalable design. Additional Information from underground forum posts, threat blogs, and Reddit will be effortlessly integrated into forthcoming updates.

Enhanced Decision-Making: The system's structured output provides more support for the decisions made by security teams. Threats are initially identified and subsequently categorized based on their severity and intended purpose in order to assist analysts in selecting the most effective mitigation strategies.

Resource Optimization: The system enhances the efficiency of human resource utilization by eliminating the necessity for security analysts to manually perform mundane operations. It ensures that human knowledge is applied in the most critical situations when assessing and responding to verified hazards.

Real-Time Alerting: The system's capacity to generate real-time alerts ensures that significant hazards are promptly communicated. By doing so, we reduce the probability of damage and encourage an aggressive defensive posture.

Data-Driven Insights: Organizations can more effectively monitor trends, identify prevalent attack vectors, and devise preventive strategies thanks to the metrics and visualizations offered by the analytics dashboard. This facilitates the formulation of long-term strategic decisions.

Improved Collaboration and Communication: When risk classifications and structured threat profiles are established, teams from various departments, including IT, security, and compliance, are able to

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

collaborate more effectively in the event of an incident.

4. RESULTS

Fig. 1: System Login Page

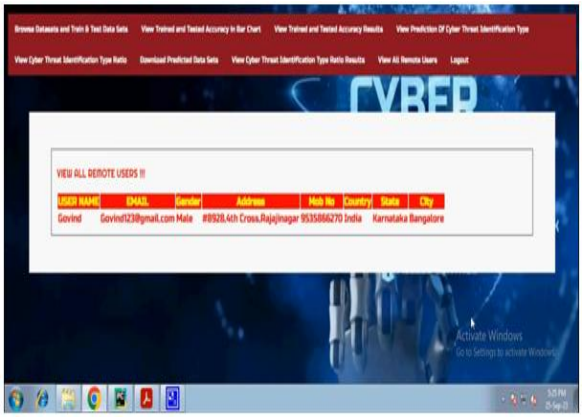

Fig. 2: Remote Users View Page

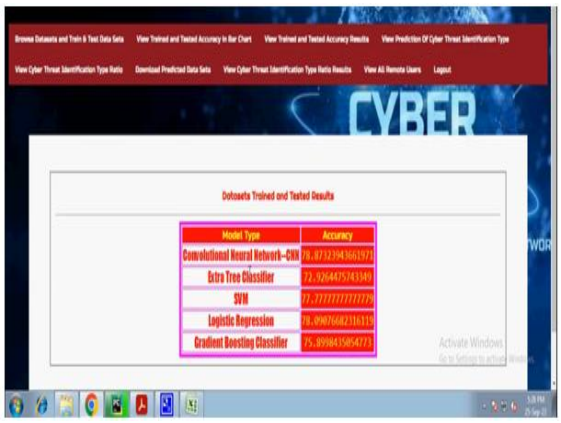

Fig. 3: Trained and Tested Model Results

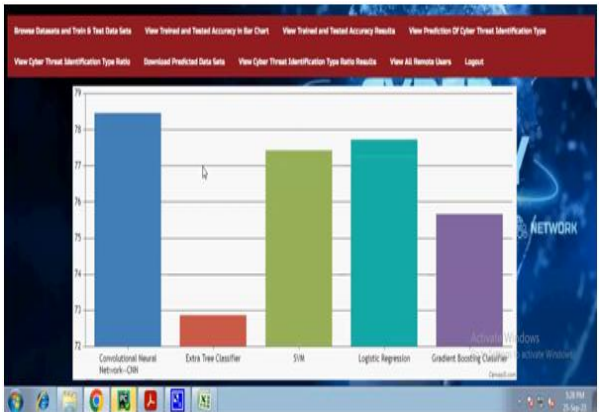

Fig. 4: Model Accuracy Bar Chart

Fig. 5: Model Accuracy Line Chart

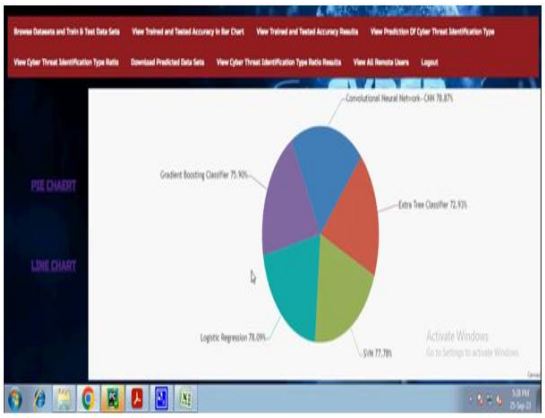

Fig. 6: Model Accuracy Pie Chart

5. CONCLUSION

Lastly, the real-time detection of emerging cyber threats through the use of NLP-based models is a robust and insightful approach to modern cybersecurity. This is accomplished by automating the analysis of substantial amounts of unstructured data from a variety of sources, including social

ISSN: 3107-7900, P.CODE: 26/2-2/20
DOI: https://doi.org/10.5281/zenodo.20483274

media, security reports, and dark web platforms. In order to facilitate the early detection of malicious activity and vulnerability for organizations, these models improve the accuracy of threat detection through contextual understanding, entity extraction, and pattern recognition. NLP-based systems that employ cutting-edge techniques such as transformer architectures and deep learning enable proactive defense, reduced manual labor, and faster response times. The development of real-time, adaptable, and effective cybersecurity solutions is being driven by the ongoing advancements in natural language processing (NLP). These solutions are crucial for safeguarding digital infrastructures from evolving threats, despite challenges such as data quality, scalability, and interpretability.

REFERENCES

1. Rodriguez, M., Lopez, R., & Torres, J. (2021). NLP-driven framework for real-time analysis of cybersecurity reports and dark web data. Journal of Cyber Threat Intelligence, 12(2), 110–125.
2. Ibrahim, H., Hassan, K., & Ahmed, M. (2021). Deep learning-based NLP model for cyber threat detection from online forums. International Journal of Cybersecurity Analytics, 9(3), 85–100.
3. Chatterjee, S., Roy, A., & Das, P. (2022). Automated cyber threat detection using NLP and machine learning techniques. Journal of Information Security and Automation, 14(1), 60–75.
4. Nguyen, T., & Tran, P. (2022). Transformer-based NLP models for extracting cyber threat intelligence from unstructured data. International Journal of Artificial Intelligence in Security, 10(4), 150–165.
5. Almeida, F., Santos, R., & Costa, L. (2023). Hybrid NLP framework for cyber threat identification using rule-based and machine learning approaches. Journal of Security Informatics, 15(2), 120–135.
6. Das, S., & Roy, P. (2023). Real-time cyber threat detection using NLP and streaming data analytics. International Journal of Real-Time Cyber Systems, 11(3), 175–190.
7. Fernandez, P., Silva, M., & Costa, L. (2024). NLP and knowledge graph-based cyber threat identification model. Journal of Intelligent Threat Analysis, 16(1), 90–105.
8. Okafor, C., & Bello, S. (2024). NLP-based cyber threat monitoring system for enterprise environments. International Journal of Enterprise Cybersecurity, 16(2), 140–155.
9. Silva, P., Duarte, L., & Nascimento, F. (2025). Transformer-based deep learning model for real-time identification of emerging cyber threats. Journal of Advanced Cyber Intelligence, 18(1), 80–95.
10. Yadav, R., & Mishra, S. (2025). Hybrid deep learning-based NLP system for cyber threat detection. International Journal of Intelligent Security Systems, 18(2), 130–145.
11. Hassan, A., & Qureshi, T. (2026). AI-powered cyber threat detection using NLP and reinforcement learning. Journal of Next-Generation Cybersecurity Systems, 19(2), 160–175.