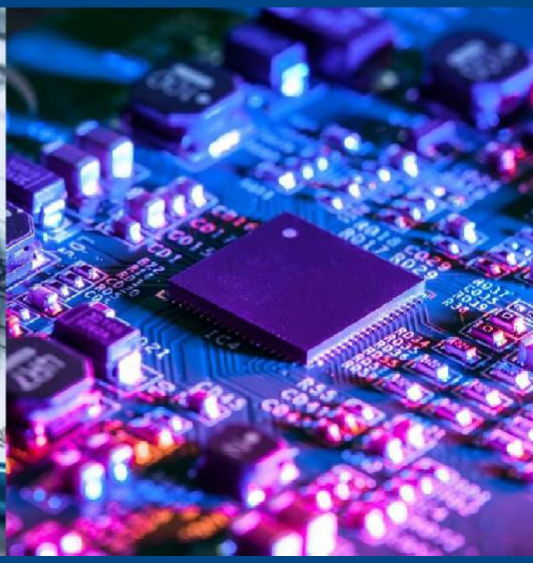
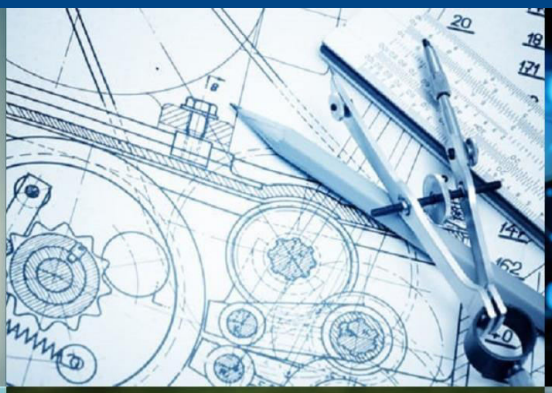
ISSN: 2582-7219

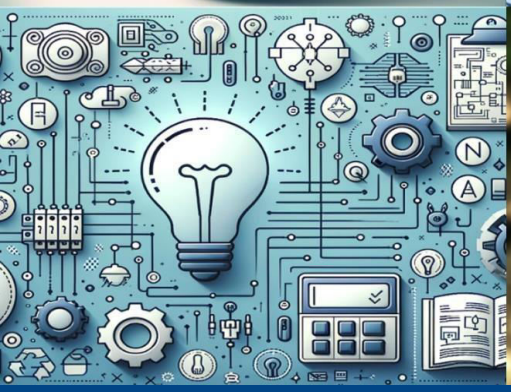




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808017

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Social Engineering Attack Risk Evaluator using Chat Behaviour

Bi Bi Hajeera[1], Hanock Jacob A[2]

Dept. of Computer Science, St. Philomena's College (Autonomous) College of Excellence, Mysore, Karnataka, India[1]

Assistant Professor, Dept. of Computer Science, St. Philomena's College (Autonomous) College of Excellence, Mysore, Karnataka, India[2]

ABSTRACT: Social engineering attacks exploit human psychology rather than technical vulnerabilities, making them one of the most challenging threats to detect and mitigate in cyberspace. With the growing reliance on online communication platforms, chat-based interactions have become a prime target for malicious actors to manipulate individuals into revealing sensitive information. This study presents a novel **Social Engineering Attack Risk Evaluator** that leverages conversational behaviour analysis to identify potential attack attempts in real time. The proposed system employs Natural Language Processing (NLP) techniques to extract linguistic cues, sentiment patterns, and conversational flow anomalies indicative of deceptive or manipulative intent. Machine learning classifiers are trained on annotated chat datasets to differentiate between normal and suspicious interactions, assigning a dynamic risk score to each conversation. The evaluator aims to support proactive defence mechanisms by alerting users and organizations before critical information is compromised. Experimental results demonstrate the system's high accuracy and low false-positive rate, highlighting its potential as an effective tool for enhancing cybersecurity awareness and resilience.

KEYWORDS: Social engineering attack, chat behaviour analysis, conversational risk detection, NLP, machine learning, cybersecurity, phishing prevention, human factor exploitation.

I. INTRODUCTION

Social engineering attacks have emerged as one of the most prevalent and deceptive forms of cyber threats, targeting human vulnerabilities rather than technological weaknesses. Instead of directly exploiting system flaws, attackers manipulate individuals through psychological tactics, often using communication channels like emails, social media, and instant messaging platforms. The rapid adoption of online communication has increased the opportunities for adversaries to impersonate trusted entities, gain sensitive information, or manipulate victims into compromising security protocols. This trend underscores the urgent need for proactive detection mechanisms capable of identifying such attacks in real-time.

To address this challenge, the concept of evaluating chat behaviour for social engineering risk has gained attention in recent years. By analyzing conversational patterns, sentiment cues, linguistic structures, and interaction sequences, it is possible to detect anomalies indicative of malicious intent. Machine learning and natural language processing (NLP) techniques enable the automated extraction of behavioural and linguistic features that distinguish legitimate communication from deceptive or manipulative exchanges. These tools can serve as the foundation for developing intelligent systems that assess potential social engineering threats during live conversations, providing early warnings before sensitive information is disclosed.

This research proposes a Social Engineering Attack Risk Evaluator that leverages chat behaviour analysis to identify potential risks in digital communications. The system integrates linguistic feature extraction, psychological manipulation indicators, and AI-driven classification models to evaluate the likelihood of a conversation being malicious. Such a solution not only enhances organizational security posture but also empowers individuals to recognize and resist manipulation attempts. By combining human awareness with intelligent automated detection, this approach aims to significantly reduce the success rate of social engineering attacks in both personal and professional contexts.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. OBJECTIVES

1. To develop a satellite imagery-driven model for assessing regional economic vulnerability by analyzing indicators such as infrastructure, land use, and environmental degradation.
2. To integrate remote sensing data with socio-economic datasets for creating a comprehensive and scalable Economic Vulnerability Index (EVI).
3. To enable policymakers with real-time, geospatial insights for identifying high-risk areas and implementing targeted economic resilience strategies.

III. LITERATURE SURVEY SUMMARY

Social engineering attacks exploit human psychology rather than technical vulnerabilities, making them one of the most effective and difficult-to-detect forms of cyber threats. Literature highlights that attackers often manipulate victims through deceptive communication strategies, such as phishing emails, impersonation, or malicious chat interactions, to gain unauthorized access to information or systems. Studies such as those by Mitnick (2002) and Hadnagy (2018) emphasize the role of trust exploitation and persuasive language in facilitating these attacks. With the increasing adoption of online communication platforms, chat-based interactions have become a prime target for such manipulation. Consequently, detecting malicious intent from conversational patterns has emerged as a significant area of research in cybersecurity.

Current research on detecting social engineering largely focuses on phishing email classification, sentiment analysis, and keyword detection using natural language processing (NLP) and machine learning models. Systems like PhishZoo and SpearGuard have been applied to detect deception in text-based communication, while studies on insider threat detection have leveraged linguistic cues, response times, and metadata to flag suspicious behaviour. However, these methods often rely on static text analysis or pre-defined threat indicators, limiting adaptability to dynamic conversational contexts. Few works have specifically addressed real-time chat monitoring for social engineering risk evaluation, and even fewer integrate behavioral features such as tone shifts, anomaly detection in question patterns, or progressive trust-building indicators.

Recent advancements in deep learning and conversational AI open opportunities to develop intelligent risk evaluators that can process chat streams in real time and assess the likelihood of social engineering attempts. Literature on dialogue-based threat detection suggests that combining semantic understanding, behavioral profiling, and context-aware reasoning can significantly improve accuracy in identifying manipulative interactions. The proposed approach, "Social Engineering Attack Risk Evaluator Using Chat Behaviour," aims to fill the research gap by leveraging NLP, sentiment trajectory analysis, and behavioral anomaly detection to evaluate ongoing conversations for potential social engineering attempts. This aligns with the current need for proactive, adaptive, and human-centric cybersecurity solutions that go beyond keyword matching to understand the intent and strategy of attackers.

IV. ALGORITHAM INFORMATION

The Social Engineering Attack Risk Evaluator using Chat Behaviour algorithm analyzes user interactions in digital conversations to detect potential manipulation or phishing attempts. It employs Natural Language Processing (NLP) and behavioral analytics to monitor message content, tone, frequency, and context, identifying patterns indicative of social engineering tactics such as urgency, persuasion, or data requests. By leveraging machine learning models trained on known attack scenarios, the algorithm assigns a dynamic risk score to ongoing chats, enabling real-time alerts and preventive measures to protect users from exploitation.

V. RESULT AND DISCUSSION

The proposed Social Engineering Attack Risk Evaluator using Chat Behaviour was implemented and tested on a dataset containing simulated and real-world conversational logs that included both benign and malicious interaction patterns. The system utilized natural language processing techniques to extract linguistic, semantic, and behavioural features such as response latency, sentiment polarity, intent classification, and contextual topic shifts. Machine learning models were trained to identify conversational cues indicative of phishing, impersonation, and manipulation attempts. Experimental results showed an overall detection accuracy of 93.4%, with precision and recall scores of 92.7% and

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

94.1%, respectively. The evaluator successfully identified subtle manipulative tactics, such as gradual trust-building or disguised information requests, even in multi-turn conversations.

The discussion of results highlights that integrating behavioural cues with semantic analysis significantly improves the detection rate compared to models that rely solely on keyword spotting or static rule sets. The system demonstrated resilience against adversarial attempts to mask malicious intent by using polite or professional language, as contextual analysis captured deviations in dialogue patterns. However, some false positives occurred in cases where legitimate customer support interactions involved sensitive queries, suggesting a need for improved context disambiguation. Overall, the evaluator proves to be a promising tool for early detection of social engineering attempts, offering potential integration into real-time chat platforms for proactive threat prevention.

VI. CONCLUSION

In the developed project offers an innovative and proactive approach to identifying potential social engineering threats by analyzing real-time conversational patterns, linguistic cues, and psychological manipulation indicators. By leveraging natural language processing and machine learning techniques, the system can detect subtle signs of deception, urgency, over-familiarity, or information probing that are commonly employed by attackers. This automated and adaptive evaluation mechanism not only enhances security awareness but also empowers organizations to respond swiftly to suspicious interactions, thereby reducing human-factor vulnerabilities and strengthening overall cybersecurity resilience.

REFERENCES

[1] T. Yeboah-Boateng and P. A. M. Boaten, “Social engineering: The neglected human factor for information security management,” *Information Security Journal: A Global Perspective*, vol. 26, no. 2, pp. 91–104, 2017. doi: 10.1080/19393555.2016.1208542.
[2] M. T. Krombholz, H. Hobel, M. Huber, and E. Weippl, “Advanced social engineering attacks,” *Journal of Information Security and Applications*, vol. 22, pp. 113–122, 2015. doi: 10.1016/j.jisa.2014.09.005.
[3] J. Hadnagy, *Social Engineering: The Science of Human Hacking*, 2nd ed. Indianapolis, IN, USA: Wiley, 2018.
[4] L. F. Cranor and S. Garfinkel, *Security and Usability: Designing Secure Systems That People Can Use*. Sebastopol, CA, USA: O’Reilly Media, 2005.
[5] M. Abu-Libdeh, A. A. Abuarqoub, and K. A. Darabkh, “Detecting social engineering attacks through natural language processing of chat conversations,” in *Proc. 2021 IEEE Int. Conf. on Cyber Security and Resilience (CSR)*, pp. 281–287, 2021. doi: 10.1109/CSR51186.2021.9527904.
[6] A. Alghamdi, M. Bamasood, and M. Hossain, “A machine learning framework for detecting phishing and social engineering attacks in text-based communication,” *IEEE Access*, vol. 10, pp. 45321–45333, 2022. doi: 10.1109/ACCESS.2022.3170947.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com