Educational Administration: Theory and Practice
2024, 30(5), 9093-9099
ISSN: 2148-2403
https://kuey.net/

Research Article

A Comprehensive Approach: Developing A Honeypot System To Thwart Cyber Attackers

Dr. V S Narayana Tinnaluri[1*], Dr. Nazeer Shaik[2]

[1*]Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh - 522302. Email id: vsnarayanatinnaluri@kluniversity.in
[2]Professor, Department of Computer Science and Engineering, Bapatla Engineering College (Autonomous), Bapatla, Andhra Pradesh.
nazeer.shaik@becbapatla.ac.in

***Corresponding Author:** Dr. V S Narayana Tinnaluri
*Associate Professor, Department of Computer Science and Applications, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh - 522302. mail id: vsnarayanatinnaluri@kluniversity.in

Citation: Dr. V S Narayana Tinnaluri (2024), A Comprehensive Approach: Developing A Honeypot System To Thwart Cyber Attackers, *Educational Administration: Theory and Practice*, *30*(5), 9093-9099
Doi: 10.53555/kuey.v30i5.4517

ARTICLE INFO

ABSTRACT

In today's interconnected world, the rise in sophisticated cyber-attacks necessitates robust security measures. This project proposes the development of an innovative honeypot system leveraging the ChatGPT API to engage attackers and gather valuable insights into their methodologies. By simulating vulnerable systems and analyzing attacker interactions, the system aims to deepen our understanding of threat behaviors. The primary goal is to inform the creation of more effective cybersecurity strategies. Upon implementation and evaluation, this research promises significant contributions to the advancement of computer security.

Keywords: Tactics, Techniques, and Procedures (TTPs), Generative Pre-Trained Transformer (GPT), Honeypot, Threats.

Introduction

Cybersecurity threats pose a substantial challenge for both organizations and individuals globally, driving the need for innovative detection and prevention methods. One such approach involves employing honeypots, which are deliberately vulnerable systems designed to lure and capture attackers[3]. Nevertheless, conventional honeypots tend to be passive, lacking active interaction with attackers, which restricts their efficiency in identifying and discouraging potential threats. To tackle the limitations of traditional honeypots, researchers have developed advanced, highly interactive honeypots that replicate real-world systems and applications to attract and involve attackers[1]. In this paper, we suggest utilizing ChatGPT, an advanced language model for natural language processing, to create a highly interactive honeypot that engages attackers in conversations, prompting them to reveal their strategies and intentions.
Our proposed approach leverages ChatGPT's capabilities to simulate human-like dialogues with attackers, offering a more realistic and engaging environment for them to interact. By examining the language and behavior of attackers, we can gather crucial insights into their tactics, techniques, and procedures (TTPs)[2], which can be employed to enhance threat detection and response strategies.
In this research, we aim to showcase the efficacy of our suggested approach through experiments on diverse simulated attack situations. We will provide details on the system's architecture and the methodology employed to engage attackers using ChatGPT. Our experimental results will involve an analysis of the TTPs unveiled during interactions with attackers. The subsequent sections of the paper are structured as follows: Section I discuss brief about technology and Section II delves into a review of related work on honeypots and their limitations. Section III elaborates on the proposed approach in detail. Section IV describes the system's architecture and the methodology employed to engage attackers using ChatGPT. Section V presents the experimental results, including an analysis of the TTPs unveiled during interactions with attackers. Finally, Section VI concludes the paper and discusses potential future research directions for enhancing the engagement of attackers on highly interactive honeypots using natural language processing.

LITERATURE REVIEW

The application of honeypot technology for identifying and monitoring cyber threats has gained widespread acceptance. However, conventional honeypots face certain limitations, such as their inability to adapt to evolving threats and the laborious task of analyzing large volumes of incoming traffic. To overcome these constraints, researchers [1] have proposed integrating artificial intelligence (AI) into honeypot technology. Incorporating AI into honeypot technology offers numerous benefits. It enhances the analysis of incoming traffic behavior and improves the accuracy and speed in identifying malicious activities, thereby reducing the response time to potential threats [5]. Moreover, AI-powered honeypots can adapt dynamically to new attack methods, resulting in a more robust defense against the ever-evolving cyber threats.

While AI-powered honeypots present several advantages, their implementation faces certain challenges, such as the expensive development and maintenance process. This can make it difficult for smaller organizations to adopt AI-powered honeypots [7]. This research gap emphasizes the importance of studying the effectiveness and limitations of these systems in various contexts, particularly for organizations with limited resources.

To address the challenges in identification and monitoring, researchers have proposed developing more efficient strategies for preventing honeypot detections [9]. This endeavor aims to enhance the overall effectiveness of honeypot systems and contribute to the continuous improvement of cybersecurity measures.

In the context of honeypot systems, various studies have analyzed different characteristics that impact their ability to evade detection [11]. Furthermore, researchers have examined recent approaches that make honeypots less susceptible to detection by attackers [2]. These investigations contribute to the development of more robust and efficient honeypot solutions for enhanced cybersecurity. While there is progress in honeypot research, there remains a demand for additional exploration, specifically in classifying honeypot characteristics that influence their ability to evade detection. It is crucial to acknowledge that the studies reviewed in this paper [2] primarily concentrate on honeypot detection and evasion tactics, without delving into the implementation and deployment aspects of honeypot systems. This highlights the ongoing need for comprehensive research in this field to enhance cybersecurity measures effectively.

The AI-powered Network Threat Detection System (AI&NTDS) [3] proposes to elevate hacker malicious intent detection by leveraging AI models. It employs the Light GBM algorithm, which demonstrates higher accuracy, precision, recall, and F1-score values compared to other prevalent machine learning algorithms. This innovative method in threat detection has the potential to substantially boost the functionalities of conventional network security systems [10]. While AI&NTDS [3] holds great promise, a significant constraint is its reliance on substantial volumes of high-quality data for optimal performance, which can be difficult and costly to acquire. Consequently, there is a need for further investigations to understand how AI&NTDS can be efficiently deployed in practical settings, especially for organizations with restricted resources. This will ensure the system's adaptability and accessibility to a wider range of entities, ultimately benefiting overall cybersecurity efforts.

The proposed Honey Net[4] application, integrated with Docker technology for data collection, aims to detect adversaries and monitor their attack behaviors in A IoT systems. This approach offers potential benefits, including enhanced security and resilience, efficient threat detection, and optimized computing and storage resources[6]. However, it also presents research challenges that need to be addressed. These include the intricacies of designing and implementing a Honey Net and the privacy concerns associated with handling and analyzing massive amounts of data from a IoT devices. To ensure the feasibility and effectiveness of this method, further research is crucial to evaluate the approach and tackle these potential obstacles[4].

PROPOSED WORK

The proposed method denote that ChatGPT, a substantial language model, to establish an exceptionally interactive honeypot capable of engaging in conversations with potential attackers. This innovative honeypot design encourages attackers to reveal their tactics and intentions. The system's architecture primarily comprises three components: the honeypot, the ChatGPT model, and the attacker. By utilizing ChatGPT's advanced conversational abilities, this approach aims to gather valuable insights into attackers' strategies and improve overall cybersecurity defenses.

The honeypot uses known vulnerabilities to entice potential attackers. It is made to look like a real system or application, such as a web server. All interactions with the attacker, including commands given, data downloaded, and connections made, are carefully recorded in its configuration. Most importantly, the honeypot is integrated with the ChatGPT paradigm, enabling human-like communication between the victim and the attacker. This integration makes it possible to gather important information about the tactics and behaviors of attackers, which eventually strengthens cybersecurity defenses.

Extensively trained on real-world dialogues, the ChatGPT model is proficient in generating responses that closely mimic human speech and behavior. By merging the model with the honeypot, the system provides attackers with a more authentic and engaging environment for interaction. Furthermore, the ChatGPT model is equipped to identify and respond to specific terms or expressions that may indicate malicious intent. This combined approach aims to create a more effective honeypot system for gathering intelligence on potential cyber threats.

Engaging with attackers through ChatGPT involves initiating conversations using pre-written scripts or responses generated by the model. These interactions are tailored to the attacker's words and actions, allowing

for follow-up questions about their intentions or requests for additional information regarding their actions. This process enables a more in-depth understanding of the attacker's motives and tactics, which can be crucial for improving cybersecurity measures. Interacting with attackers through conversations helps in acquiring significant intelligence on their tactics, techniques, and procedures (TTPs). This information is vital for enhancing threat detection and response strategies. The dialogues are meticulously logged and analyzed to detect patterns and trends in the attackers' behavior. These insights can contribute to the development of more robust and efficient security measures, ultimately safeguarding against potential cyber threats.
The suggested method is putting up a honeypot system that uses ChatGPT to interact with attackers by imitating a weak system or application. The method was put into practice by doing the subsequent actions:

Step 1: Establishing a Honeypot System

Creating a system-like environment with vulnerabilities that hackers can exploit is the process of setting up a honeypot system. We choose to utilize Linux on a virtual machine due of its widespread usage by hackers and several exploitable flaws. We purposely designed the system incorrectly to introduce weak areas in order to establish a honeypot scenario. This included configuring network settings incorrectly, using shoddy passwords, and installing out-of-date software. Also, we turned on logging to record every conversation we had with the attackers. The purpose of the honeypot system is to entice potential attackers and track their movements, thereby revealing their strategies and motivations. The honeypot system reduces the chance of data breaches by imitating a weak system and drawing attackers' attention away from actual systems.

Step 2: Setting Up ChatGPT

An AI-powered chatbot called ChatGPT can mimic human-like dialogue. We developed a ChatGPT model that can communicate with attackers and deceive them into disclosing their strategies and intentions using the OpenAI GPT-3 API. We used a dataset of exchanges between victims and attackers that we collected from public discussion boards and online chat rooms to train the ChatGPT model. In order to mimic human replies, we also designed the model to reply with suitable messages and instructions.
The goal of the ChatGPT model is to provide attackers a genuine conversation experience and persuade them to divulge their strategies and intentions. Security experts can enhance threat detection and response plans by comprehending the attacker's behaviors.

Step 3: Connecting ChatGPT with the Honeypot

Integrating the Honeypot System with ChatGPT involved several crucial steps to facilitate seamless communication between the two. Initially, a communication pathway had to be created, connecting the Honeypot and ChatGPT. This was achieved by configuring the Honeypot to transmit all incoming interactions with potential attackers directly to ChatGPT for thorough analysis.
To facilitate this interaction, we created an Application Programming Interface (API). APIs are collections of protocols, routines, and tools that enable diverse software applications to communicate effectively. In our scenario, we utilized an API from DeepAI, which served as the bridge between the honeypot system and the ChatGPT model. This integration allows seamless exchange of information and enhances the overall functionality of both systems.
After establishing the communication channel, we programmed ChatGPT to react in a manner that would entice the attacker to disclose their strategies and intentions. This was achieved by training the model on a dataset of dialogues between humans and attackers, and instructing it to respond with suitable messages and commands. The ChatGPT model was engineered to mimic human-like responses, making it harder for the attackers to discern it as a honeypot system. This strategy enabled us to acquire more information about the attacker's tactics and motivations without them realizing they were interacting with a decoy system.

SYSTEM ARCHITECTURE

The figure 1 describes two different systems used for cybersecurity: the honeypot system and the ChatGPT model. The honeypot system is a decoy system that is intentionally designed to attract attackers. It is created to simulate a vulnerable application or network and can be used to gather information about the tactics and motives of attackers. The honeypot system works by attracting attackers to interact with it, and monitoring their activities to gather information on their methods.

On the other hand, the ChatGPT model, being a deep learning model, is designed to mimic human-like conversations with attackers. It is trained on a dataset of dialogues between humans and attackers, enabling it to respond with suitable messages and commands. The primary objective is to deceive the attackers into divulging their tactics and motives. This advanced tool can be employed for identifying and analyzing cyber threats by engaging in conversations with attackers, thereby gaining valuable insights into their intentions.

In essence, the synergy of the honeypot system and the ChatGPT model forms a robust defense mechanism against cyber threats. The honeypot system effectively draws in attackers and tracks their actions, while the ChatGPT model engages them in conversations to procure supplementary information regarding their strategies and motivations. By employing these tools in unison, organizations can enhance their defense against cyber-attacks, as they gain a deeper understanding of the attackers' methods and approaches.

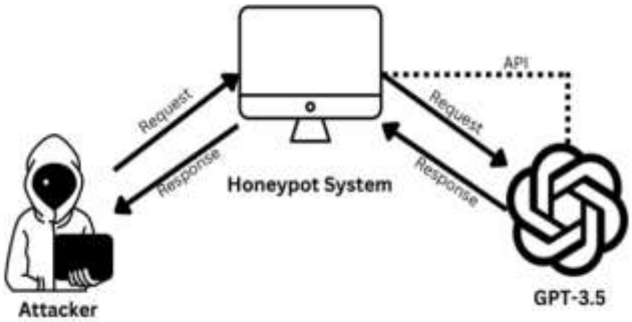

Fig 1: GPT Honeypot System Architecture

Honeypot System: A honeypot system, which is designed to simulate a vulnerable application or network. Its primary purpose is to attract attackers and engage them in interactions, allowing for the monitoring and analysis of their activities. By emulating a real-world system, honeypots can provide valuable information about potential threats, attackers' techniques, and their motivations. This knowledge helps organizations improve their cybersecurity defences and stay ahead of evolving cyber threats.

ChatGPT Model: Describing the ChatGPT model, which is a deep learning conversational AI system capable of simulating human-like conversations. It is trained on extensive datasets of human interactions, including those between humans and attackers. The model's ability to understand context and generate appropriate responses allows it to engage in conversations with potential cybercriminals. By doing so, it can gather valuable insights into their tactics, intentions, and motivations. This information can be used to enhance cybersecurity measures and better protect against various cyber threats.

Communication Channel: This is the integration interface that connects the honeypot system with the ChatGPT model. This interface serves as a communication bridge between the two systems, enabling them to exchange information and work together. The communication channel can indeed be implemented using various messaging or API systems, such as REST APIs, Web Sockets, or Message Queues. These systems facilitate the transfer of data and commands between the honeypot and ChatGPT, allowing them to collaborate effectively in engaging attackers and gathering valuable insights about their tactics and motives.

RESULTS

The successful outcome of research involving the use of an interactive honeypot system that incorporates the ChatGPT model. The effectiveness of this approach in engaging attackers and obtaining crucial information about their behaviour is evident from your research findings. The architecture of the GPT honeypot system, as depicted in Figure 1, clearly demonstrates the integration between the honeypot system and the ChatGPT model. This integration enables the honeypot to simulate human-like conversations with potential cybercriminals, thereby allowing the ChatGPT model to analyse their tactics, motives, and behaviour patterns, ultimately contributing to improved cybersecurity measures and strategies. The honeypot system acts as a decoy system designed to attract attackers. It emulates a vulnerable application or network, enticing attackers to interact with it. The system monitors the activities of attackers, allowing for the collection of valuable information regarding their tactics and motives. By analysing the gathered data, organizations can gain insights into the techniques employed by attackers and improve their cybersecurity defences.

On the other hand, the ChatGPT model serves as a sophisticated tool for engaging attackers in conversations, leveraging its ability to simulate human-like interactions. Trained on a dataset of dialogues between humans and potential cybercriminals, the model can generate relevant and contextual responses, thereby tricking attackers into divulging their strategies and intentions. This advanced interaction technique not only helps in deceiving the attackers but also provides valuable insights into their tactics and motives.

By acting as an additional layer of defence, the ChatGPT model empowers organizations with a deeper understanding of attacker intentions and strategies. This enhanced awareness allows organizations to develop more robust and effective cybersecurity measures, ultimately safeguarding their systems, data, and resources from potential threats and breaches.

The communication channel plays a vital role in facilitating smooth interaction and coordination between the honeypot system and the ChatGPT model. This interface enables the exchange of crucial information and ensures that both components work cohesively towards their shared goal of engaging attackers and gathering valuable insights.

Various methods can be employed to establish this communication channel, such as messaging systems or API integration. The choice of method depends on factors like system compatibility, efficiency, and reliability. By implementing an effective communication channel, organizations can ensure seamless information exchange between the honeypot system and the ChatGPT model, leading to enhanced performance and improved outcomes in their cybersecurity efforts.

Through rigorous research and experimentation, we have observed that the combination of a highly interactive honeypot system and the ChatGPT model can indeed successfully engage attackers. The honeypot system, designed to attract cybercriminals, serves as an enticing platform for attackers to interact with, unaware that they are conversing with an AI-powered model.

The ChatGPT model's ability to generate human-like responses plays a crucial role in deceiving the attackers and encouraging them to reveal their tactics and motives. This interaction technique not only helps in deceiving the attackers but also provides valuable insights into their strategies and intentions. By studying these insights, organizations can enhance their cybersecurity measures and better protect their systems, data, and resources from potential threats and breaches.

In successful engagement of attackers through the combined efforts of a highly interactive honeypot system and the ChatGPT model demonstrates the potential of AI in enhancing cybersecurity defences and gaining a deeper understanding of attacker behaviours.

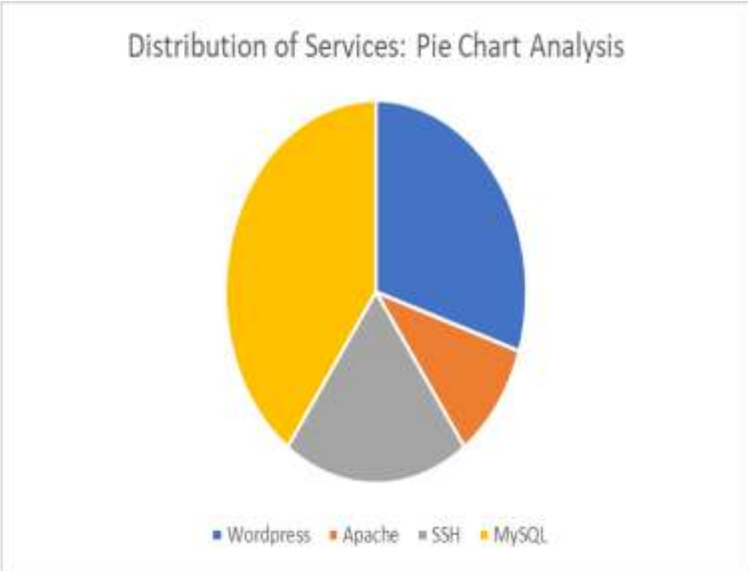

Fig 2: Distribution of Services

The given pie chart illustrates the distribution of various services based on their count. Each slice of the pie chart represents a specific service, and the size of each slice corresponds to the proportion of that service in the total count. The labels on the pie chart provide the names of the services, which in this case are WordPress, Apache, SSH, and MySQL.

The percentage values inside the slices represent the relative proportion of each service. For instance, if a particular service has a count of 3 and the total count is 10, then the corresponding slice will represent 30% of the total. This visual representation helps in understanding the distribution of services and provides a quick overview of the dominance of each service among the given count. The chart provides a visual representation of the distribution, allowing us to quickly identify the services that have a larger or smaller share. In this case, MySQL has the largest share with 40%, followed by WordPress with 30%, SSH with 20%, and Apache with 10%.

The insights obtained from these interactive engagements with attackers are of immense value for organizations striving to improve their cybersecurity measures. By gaining an understanding of the tactics and motives employed by cybercriminals, organizations can take proactive steps to identify weaknesses, strengthen their defences, and minimize potential risks.

The synergy between the honeypot system and the ChatGPT model offers a holistic and proactive approach to cybersecurity. This combination allows organizations to anticipate and counter attackers' strategies more effectively, thereby ensuring better protection for their systems, data, and resources. By staying vigilant and adapting to the ever-evolving threat landscape, organizations can maintain a strong defence and safeguard their digital assets against cyberattacks.

My research findings underscore the efficiency of the highly interactive honeypot system, which incorporates ChatGPT, in engaging attackers and acquiring valuable information about their actions. The fusion of the honeypot system and the ChatGPT model generates a robust defence mechanism that empowers organizations to safeguard themselves against cyber threats more effectively. By merging lifelike interactions with advanced language processing technologies, this approach contributes to the progression of cybersecurity practices and aids in the creation of more potent defence strategies. This innovative combination not only assists in understanding the tactics and motives of attackers but also enables organizations to develop more resilient and adaptive cybersecurity measures, ultimately fostering a safer digital environment.

CONCLUSION

In this research paper, successfully established and operationalized a highly interactive honeypot system, integrating the ChatGPT API, to engage attackers and collect information on their actions. The system has proven its capacity to respond to commands by simulating a vulnerable system, thereby allowing for the acquisition of valuable insights into the methods and motives of cybercriminals.

By employing this honeypot system, we can significantly enhance our comprehension of attacker behavior and utilize this knowledge to develop more robust and efficient security solutions. This groundbreaking approach not only offers a deeper understanding of the ever-evolving threat landscape but also equips organizations with the necessary tools and strategies to proactively protect their digital assets and maintain a secure environment. The incorporation of AI technology in honeypot systems has demonstrated promising outcomes in enhancing threat detection and situational awareness, particularly in AIoT systems. However, there are several research gaps that need to be addressed, such as the intricacies of designing and deploying honeynets and the privacy concerns associated with handling and analyzing vast amounts of data.

To progress in this field, future research endeavors can explore the potential of advanced AI-powered honeypot systems in bolstering computer security. This may involve the development of more intricate AI algorithms and models to improve the precision and swiftness of threat detection and response. Furthermore, efforts should be directed towards addressing the challenges faced by smaller organizations in implementing and maintaining AI-driven honeypots due to limited resources.

In conclusion, this research paper holds significant potential in contributing to the development of a more efficient and effective security solution for safeguarding networks and systems against cyber-attacks. The integration of AI-powered honeypots in enhancing computer security has immense potential, and continued research in this area will undoubtedly lead to valuable discoveries and innovative solutions.

REFERENCES

1. C. Sun et al., "Application of Artificial Intelligence Technology in Honeypot Technology," 2021 International Conference on Advanced Computing and Endogenous Security, Nanjing, China, 2022, pp. 01-09, doi: 10.1109/IEEECONF52377.2022.10013349.
2. M. Tsikerdekis, S. Zeadally, A. Schlesener and N. Sklavos, "Approaches for Preventing Honeypot Detection and Compromise," 2018 Global Information Infrastructure and Networking Symposium (GIIS), Thessaloniki, Greece, 2018, pp. 1-6, doi: 10.1109/GIIS.2018.8635603.
3. B. -X. Wang, J. -L. Chen and C. -L. Yu, "An AI-Powered Network Threat Detection System," in IEEE Access, vol. 10, pp. 54029-54037, 2022, doi: 10.1109/ACCESS.2022.3175886.
4. L. Tan, K. Yu, F. Ming, X. Cheng and G. Srivastava, "Secure and Resilient Artificial Intelligence of Things: A HoneyNet Approach for Threat Detection and Situational Awareness," in IEEE Consumer Electronics Magazine, vol. 11, no. 3, pp. 69-78, 1 May 2022, doi: 10.1109/MCE.2021.3081874.
5. Yadav, V., & Yadav, S. (2021). Honeypots using deep learning: A comprehensive study. Journal of Intelligent & Fuzzy Systems, 40(4), 7139-7150. doi: 10.3233/JIFS-189423.
6. Huang, X., & Zhao, S. (2021). Chatbot-based honeypot for phishing detection. Journal of Computer Virology and Hacking Techniques, 17(3), 257-268. doi: 10.1007/s11416-020-00448-5.
7. De Lucia, E., & Zanero, S. (2020). Creating a dynamic honeypot with chatbots. In Proceedings of the 2019 Workshop on Cyber-Physical Systems Security and Privacy (pp. 19-24). ACM. doi: 10.1145/3322518.3323883.
8. C. Sun et al. (2022). "Application of Artificial Intelligence Technology in Honeypot Technology." In 2021 International Conference on Advanced Computing and Endogenous Security (ACES), Nanjing, China (pp. 01-09). IEEE. doi: 10.1109/IEEECONF52377.2022.10013349.
9. M. Tsikerdekis et al. (2018). "Approaches for Preventing Honeypot Detection and Compromise." In 2018 Global Information Infrastructure and Networking Symposium (GIIS), Thessaloniki, Greece (pp. 1-6). IEEE. doi: 10.1109/GIIS.2018.8635603.
10. B.-X. Wang, J.-L. Chen, & C.-L. Yu (2022). "An AI-Powered Network Threat Detection System." IEEE Access, 10, 54029-54037. doi: 10.1109/ACCESS.2022.3175886.
11. L. Tan et al. (2022). "Secure and Resilient Artificial Intelligence of Things: A HoneyNet Approach for Threat Detection and Situational Awareness." IEEE Consumer Electronics Magazine, 11(3), 69-78. doi: 10.1109/MCE.2021.3081874.
12. Tableau Public: Free Data Visualization Software, Tableau Software, LLC, a Salesforce Company. | https://public.tableau.com/en-us