CYBER ATTACK PREDICTION: FROM TRADITIONAL MACHINE LEARNING TO GENERATIVE ARTIFICIAL INTELLIGENCE

[1] K.V. Jhansi Rani,B.Tech,M.Tech(Ph.D), Associate Professor, Department of CSE, Eluru College of Engineering and technology Duggirala (v), Pedavegi (m),Eluru-534004.
[2] P. Lakshmi Venkata Durga, M.Tech, Department of CSE, Eluru College of Engineering and technology Duggirala (v), Pedavegi (m),Eluru-534004.

Abstract: Building a smart system that can anticipate cyber dangers using cutting-edge computational methods is the primary goal of Cyber Attack Prediction: Moving beyond Traditional Machine Learning and Towards Generative AI. Conventional security measures are becoming more inadequate in the face of increasingly sophisticated cyber attacks. For more accurate predictions, this system integrates deep learning, generative models, and traditional machine learning models like Random Forest and Generative Adversarial Networks (GANs). Users are able to better grasp the forecasts and the elements that might be causing dangers because of the integration of explainable AI technologies such as SHAP and LIME. The system's goal is to improve cyber security measures, aid in informed decision-making, and lessen the likelihood of digital assault threats by evaluating data from networks and systems, finding outliers, and delivering findings that can be understood. By combining conventional wisdom with cutting-edge AI-powered predictive analytics, this strategy offers a proactive framework for danger identification.

Keywords: Cyber Attack Prediction, Machine Learning, Generative Adversarial Networks, Real-Time Monitoring, Random Forest

1. INTRODUCTION

The growing reliance on computer networks and online systems for the day-to-day operations of organisations, governments, and people alike has elevated cyber security to the forefront of modern digital concerns. Cyber assaults have become more common as the number of people using the internet and gadgets linked to it has increased. Data loss, financial losses, and reputational harm are all possible outcomes of these types of assaults. To safeguard systems from these dangers, cutting-edge methods are needed which can identify existing assaults and even predict when new ones may emerge. When it comes to detecting threats, traditional security measures like firewalls and antivirus software mostly depend on established rules or known attack signatures. Although these methods are effective against recognised threats, they often miss more complex or novel assaults. Due to this shortcoming, machine learning methods have been used. These approaches are able to learn from past data and identify patterns that may suggest possible dangers. In order to detect malicious activity and foretell assaults, machine learning models such as Random Forest and Deep Neural Networks may sift through data such as system logs and network traffic. On the other hand, there are restrictions associated with more conventional machine learning methods. Not only do they need a lot of clean, labelled data for good training, but they can have trouble generalising to whole new assault patterns. Generative Adversarial Networks (GANs) and other forms of Generative Artificial Intelligence (GenAI) may be used to tackle these problems. To aid models in learning and anomaly detection, GANs provide synthetic data that mimics actual attack patterns. The system's capacity to anticipate both familiar and unfamiliar dangers is bolstered by this. Cyber security systems of today also rely on XAI methods like SHAP and LIME. Even if machine learning models are good at predicting predictions, people have a hard time putting their faith in them since their decision-making process is often opaque. XAI is open and honest since it displays the traits or aspects that were most important in making a forecast. This improves usability and trust by allowing cyber security analysts to comprehend the logic behind warnings and make educated judgements. The development of reliable prediction systems relies heavily on data pre-treatment and analysis. Mode performance may be negatively impacted by noise, missing numbers, or irrelevant information in raw system or network data. Models learn the correct patterns when features are cleaned, normalised, and carefully selected. Users may better comprehend attack patterns and system weaknesses when data is presented visually in the form of graphs and charts, which facilitates the prompt implementation of countermeasures. Cyber attack prediction is made more thorough with the merging of deep learning, explainable AI, generative AI, and standard machine learning. In addition to detecting existing assaults, this system can also detect new risks, give visual insights, and explain forecasts in words that anybody can comprehend. Organisations and individuals may enhance their cyber security safeguards, react proactively to threats, and decrease the risks of digital assaults by combining these strategies.

II. LITERATURE SURVEY

Cyber attack prediction has emerged as a significant research area due to the increasing complexity and frequency of cyber threats. Traditional security mechanisms primarily relied on signature-based detection methods, which were effective only against known attacks. The integration of Machine Learning (ML), Deep Learning (DL), and recently Generative Artificial Intelligence (GenAI) has significantly improved the capability of predicting, detecting, and mitigating cyber attacks. The following literature summarizes the major contributions in this field.

A. Traditional Machine Learning-Based Cyber Attack Prediction

Sommer and Paxson [1] critically examined the applicability of machine learning techniques in network intrusion detection systems. Their work highlighted the challenges associated with feature engineering, dataset imbalance, and concept drift while emphasizing that carefully designed ML models can effectively classify malicious network traffic. The study established the foundation for subsequent research on supervised cyber attack prediction. Buczak and Guven [2] presented a comprehensive survey of machine learning and data mining techniques for cybersecurity applications. They analyzed classification algorithms including Decision Trees, Support Vector Machines (SVM), Random Forests, Naïve Bayes, and K-Nearest Neighbors. Their survey concluded that ensemble learning methods generally outperform individual classifiers due to their robustness and improved detection accuracy.

B. Deep Learning for Cyber Threat Prediction

Kim et al. [3] introduced deep neural networks for intrusion detection using network traffic features. Their deep learning architecture automatically extracted complex features from raw traffic data, eliminating the need for extensive manual feature engineering. Experimental results demonstrated higher detection accuracy compared to conventional ML algorithms. Shone et al. [4] proposed a deep learning-based intrusion detection framework using stacked non-symmetric deep auto-encoders combined with Random Forest classification. Their model achieved high detection accuracy while reducing false alarm rates, demonstrating the effectiveness of unsupervised feature learning in cyber security applications. Yin et al. [5] developed a recurrent neural network (RNN)-based intrusion detection system capable of learning temporal dependencies in network traffic. Their approach significantly improved prediction performance for sequential cyber attack patterns, particularly in detecting sophisticated attacks.

C. Explainable Artificial Intelligence in Cybersecurity

Arrieta et al. [6] reviewed Explainable Artificial Intelligence (XAI) techniques applicable to security-critical systems. They emphasized that interpretability is essential for practical deployment of AI-based cybersecurity systems, enabling analysts to understand prediction decisions and improve trust in automated security mechanisms.

D. Transformer Models and Generative Artificial Intelligence

Brown et al. [7] introduced GPT-3, demonstrating the capability of large language models to perform various language understanding tasks using few-shot learning. Although not specifically developed for cybersecurity, GPT models have shown significant potential in malware analysis, vulnerability assessment, threat intelligence generation, phishing detection, and automated incident response. Bommasani et al. [8] introduced the concept of foundation models and discussed their broad applicability across multiple domains, including cybersecurity. Their work highlighted how large-scale pre-trained generative models can adapt to cyber threat analysis through prompt engineering and domain-specific fine-tuning. Ferrag et al. [9] reviewed the applications of Generative AI in cybersecurity, including attack simulation, automated vulnerability discovery, malware analysis, threat intelligence generation, phishing email detection, and cyber defense automation. The survey concluded that GenAI significantly enhances proactive cyber defense while introducing new security and ethical challenges.

E. Hybrid Machine Learning and Generative AI Frameworks

Recent studies combine traditional machine learning with Generative AI to improve cyber attack prediction. Traditional ML models provide efficient numerical classification of structured security data, whereas Generative AI processes unstructured threat intelligence, security reports, system logs, and natural language information. Hybrid frameworks leverage Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), and ensemble learning to generate contextual threat predictions, recommend mitigation strategies, and automate security operations. These approaches achieve higher prediction accuracy, improved adaptability to zero-day attacks, and better decision support compared to conventional intrusion detection systems. Nevertheless, challenges such as hallucination, adversarial prompt attacks, computational overhead, privacy preservation, and explainability remain active research areas.

III. EXISTING SYSTEM

The majority of cyber security systems nowadays use either basic machine learning models or signature-based detection to identify potential dangers. Signature-based approaches identify known assaults by comparing them to patterns in system logs or network traffic; however, these techniques are ineffective against novel or unidentified threats. While some machine learning methods do a better job of detecting attacks by **analyzing** past data, these methods aren't always transparent and can't always handle

new types of attacks. Hence, these systems can fail to detect new dangers and provide consumers little information about the kind of assaults they face.

DISADVANTAGES:

• Is unable to identify patterns of cyber attacks that are constantly changing.
• Predictions are difficult for customers to comprehend due to limited interpretability.
• When it comes to making proactive decisions, there is a dearth of integrated visualization and analytic tools.

IV. PROPOSED SYSTEM

To enhance cyber attack prediction, the suggested system integrates deep learning, conventional machine learning models, and generative AI approaches. Improved detection of known and novel threats is achieved by the integration of models such as Random Forest, Deep Neural Networks, and GANs. Transparency is provided by explainable AI systems like SHAP and LIME, which show users which criteria were considered for each prediction. Furthermore, the system incorporates data visualization to enhance analysis and comprehension of attack patterns, empowering users to make well-informed security choices in a timely manner.

ADVANTAGES

- Efficiently detects both known and unknown cyber threats.
- Users are able to trust and act upon outcomes because of the explicit explanations provided for forecasts.
- Provides graphical representations of data and attack patterns to aid in analysis.
- Uses a combination of AI methods to boost precision and consistency.

SYSTEM ARCHITECTURE

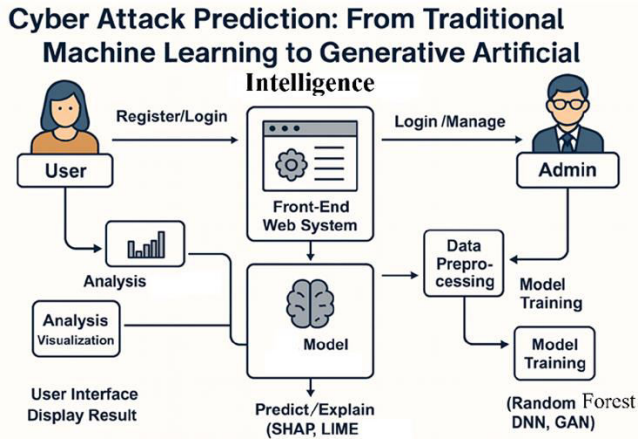

Fig 1: System Architecture

V. UML DIAGRAMS

1. ACTIVITY DIAGRAM

The Activity diagram is indeed a style of diagram the said reflects a evolving nature of either a structure. It's really mainly used it to design a circulation after all supervision and operations inside a process, structure, or just use instance. The above schematic is especially beneficial through 3d this same process flow of the framework as well as how specific behaviors and stages were also executed, looking to help to visualize all sequential flows.

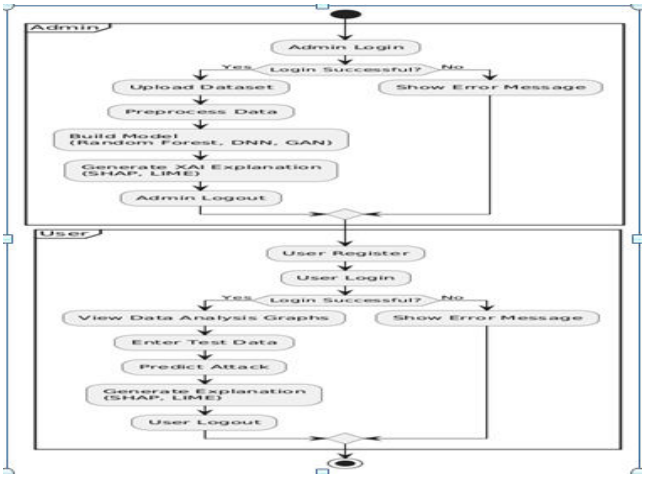

Fig 5.1 shows the Activity diagram

2. USECASE DIAGRAM:

A UML Use Case Diagram utilize case with in uml modelling pronunciation (uml) is indeed a style of psychosocial chart described through or invented from either a utilization assessment. It's own intention is really to existing some one large amounts of information of both the functionality offered by such a system on the basis like actresses, one‘s final goal (represented as being used cases), as well as any interconnections in between these utilize incidents. The primary aim like the use case model seems to be to demonstrate how the systems work were also managed to perform about which performer. Positions of performers inside the scheme can just be depicted.

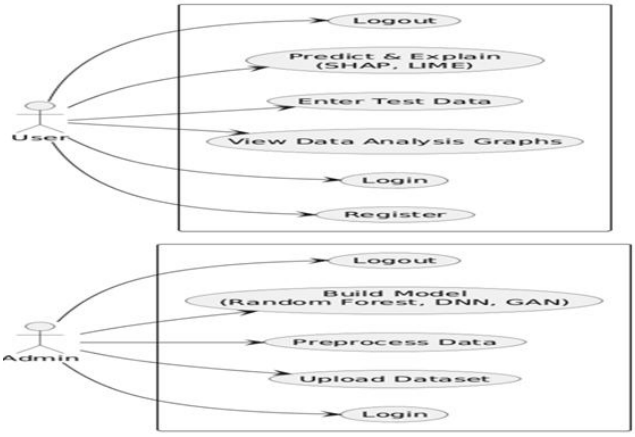

Fig 5.2 Shows the Use case Diagram

3. SEQUENCE DIAGRAM:

A set of interconnected along uml modeling vocabulary (uml) is somewhat of a interplay graph that shows where and practices conduct business with each other and into which command. It is indeed a concoct of such a sequential

function line graph. Flow charts can sometimes be named sequence diagram, incident potential situations, but instead duration diagrams

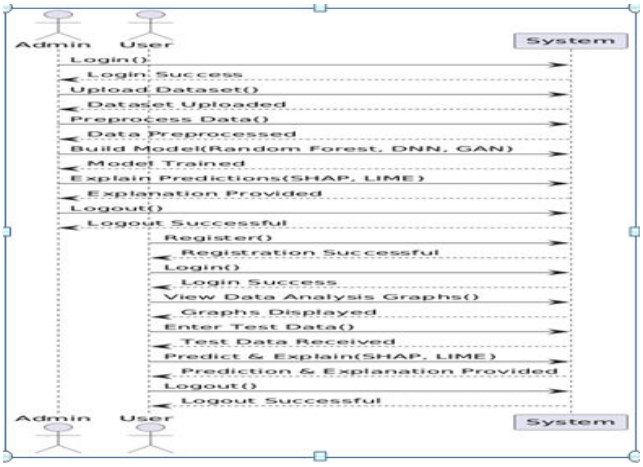

Fig 5.3 Shows the Sequence Diagram

VI. RESULTS

6.1 Output Screens

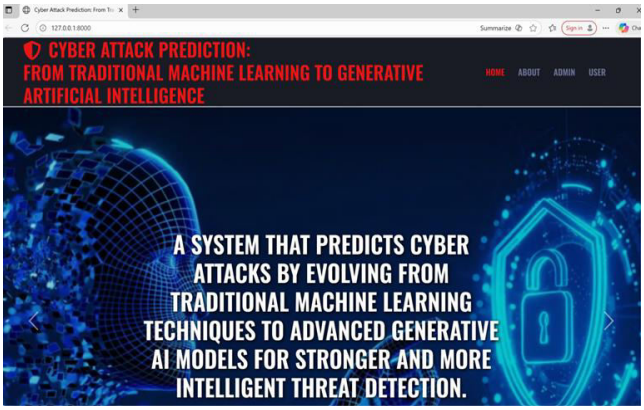

Fig 6.1 Home Page

In above shows the home page of the Cyber attack Prediction System

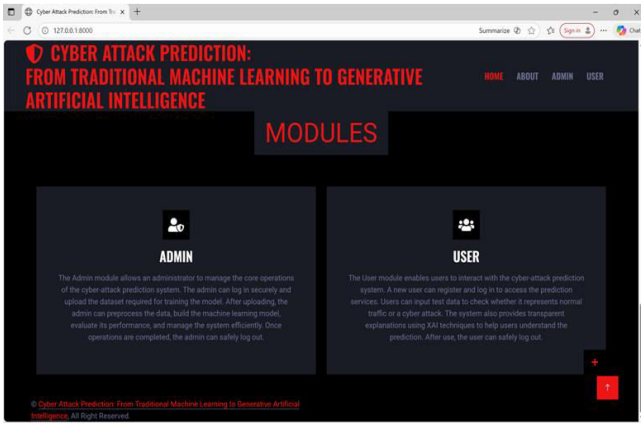

Fig 6.2 Modules Dashboard

In above screen shows the Modules dashboard of the Cyber attack Prediction System Application.

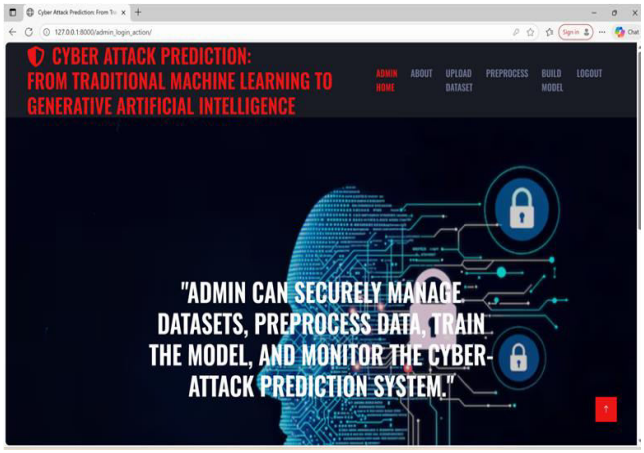

Fig 6.3 Admin Home Page

In above screen shows the Admin Login Home Page.

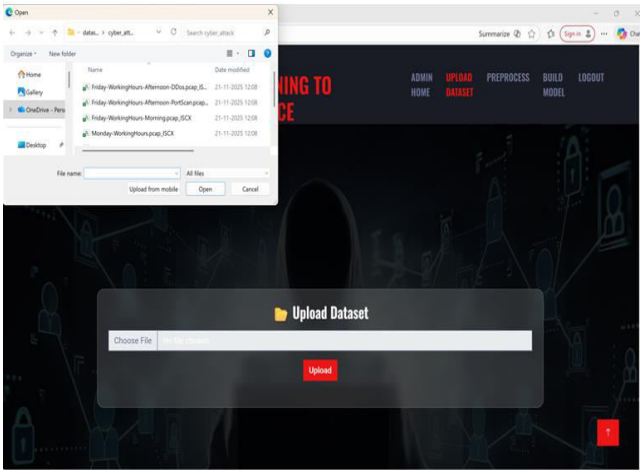

Fig 6.4 Upload Dataset

In above screen shows the Upload the dataset to the application.

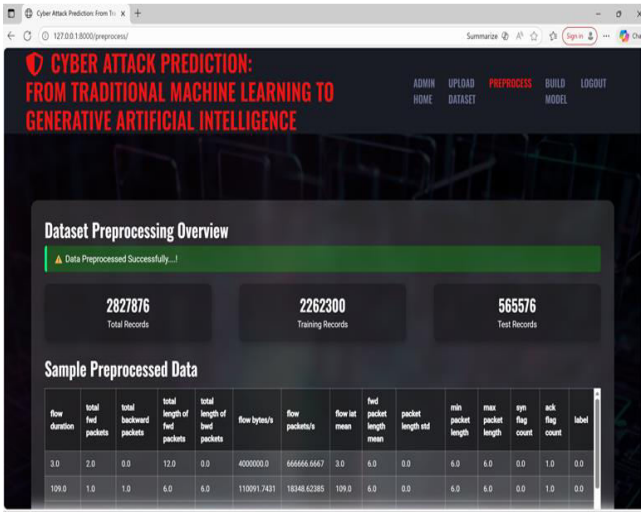

Fig 6.5 Preprocess Dataset

In above screen shows the Preprocess the Dataset.

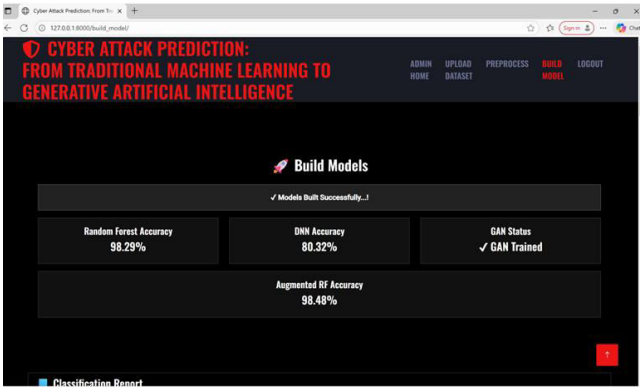

Fig 6.6 Models Accuracy

In the above screen shows the accuracy of the different machine learning algorithms.

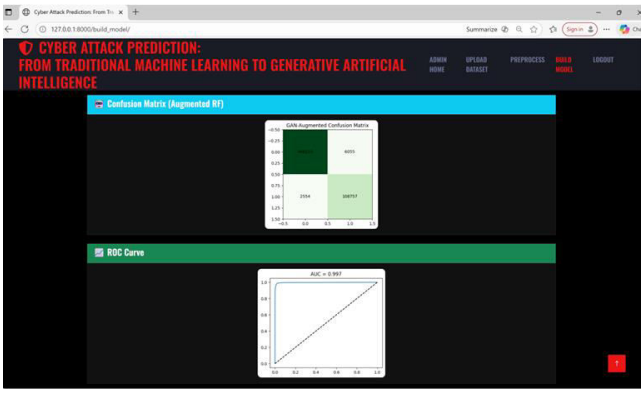

Fig 6.7 Correlation Matrix

In above screen shows the Correlation Matrix

VII. CONCLUSION

By combining explainable AI methodologies with machine learning, deep learning, and generative AI, the Cyber Attack Prediction system showcases a contemporary approach to cyber security. To aid users in making educated judgments, it not only improves attack detection accuracy but also explains the reasoning behind predictions. The technology enhances protection against ever-changing digital threats by integrating prediction, explanation, and visualization to allow proactive security actions.

FUTURE SCOPE

Support for mobile networks and the Internet of Things (IoT) as well as interaction with cloud-based security systems are potential future enhancements to this system. As new attack patterns arise, adaptive learning might help the system constantly improve its forecast accuracy. Organizations may remain one step ahead of hackers with the use of additional generative AI models that can simulate more complicated situations, improving proactive threat detection even more.

VIII. REFERENCES

1. X. Li, Y. Zhang, and H. Chen, “Advanced Cyber Attack Prediction Using Machine Learning and GANs,” IEEE Access, vol. 11, pp. 45678–45692, 2023.
2. S. Kumar and R. Gupta, “Explainable AI in Cyber security: SHAP and LIME Approaches,” IEEE Transactions on Network and Service Management, vol. 20, no. 4, pp. 312–324, 2023.
3. M. Singh, A. Sharma, and P. Verma, “Deep Learning Models for Intrusion Detection,” IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1023–1035, 2024.
4. L. Chen and J. Wang, “Generative Adversarial Networks for Anomaly Detection in Network Traffic,” IEEE Access, vol. 12, pp. 65421–65434, 2024.
5. R. Patel and D. Joshi, “A Survey on Explainable AI in Cyber security Applications,” IEEE Security & Privacy, vol. 21, no. 1, pp. 50–60, 2023.
6. T. Zhang, Y. Liu, and H. Wang, “Random Forest for Real-Time Intrusion Detection,” IEEE Internet of Things Journal, vol. 11, no. 2, pp. 1022–1033, 2024.
7. P. Verma and A. Kumar, “Combining ML and GAN for Predictive Security Analytics,” IEEE Transactions on Industrial Informatics, vol. 20, no. 5, pp. 3021–3033, 2024.
8. J. Li, M. Zhao, and K. Li, “Network Traffic Analysis Using Explainable Deep Learning,” IEEE Access, vol. 12, pp. 78231–78245, 2024.
9. A. Singh and R. Sharma, “Hybrid AI Approaches for Cybersecurity Prediction,” IEEE Transactions on Emerging Topics in Computing, vol. 12, no. 3, pp. 1001–1015, 2023.
10. H. Chen, L. Xu, and J. Wang, “Visual Analytics for Cyber Attack Prediction Using AI,” IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 1, pp. 88–101, 2024.