ISSN No:-2456-2165

Exploring the Fusion of Graph Theory and Diverse Machine Learning Models in Evaluating Cybersecurity Risk

[1]Clive Ebomagune Asuai,
Computer Programmer / System Analyst,
[1]ICT Department, Delta State Polytechnic, Otefe-Oghara

[2]Gideon Yuniyus Giroh
Post Graduate Student
[2]Modibbo Adama University, Yola

Abstract:- The frequency and severity of cyber-attacks have surged, causing detrimental impacts on businesses and their operations. To counter the ever-evolving cyber threats, there's a growing need for robust risk assessment systems capable of effectively pinpointing and mitigating potential vulnerabilities. This paper introduces an innovative risk assessment technique rooted in both Machine Learning and graph theory, which offers a method to evaluate and foresee companies' susceptibility to cybersecurity threats. In pursuit of this objective, four Machine Learning algorithms (Random Forest, AdaBoost, XGBoost, Multi-Layer Perceptron (MLP)) will be employed, trained, and assessed using the UNSW-NB15 dataset that has a hybrid of real modern normal activities and synthetic contemporary attack behaviours..The findings indicate that the Multilayer Perceptron (MLP) performs better than other classifiers, achieving an accuracy of 98.2%.. By harnessing the capabilities of data-derived insights and intricate network analysis, this groundbreaking approach aims to equip organizations with a comprehensive and forward-looking cybersecurity defense strategy.

Keywords:- *Cyber-Attacks, Risk Assesment, Graph Theory, Multi-Layer Perceptron, AdaBoost, Random Forest, XGBoost*

I. INTRODUCTION

In recent decades, the rapid growth of information technology has led to a notable surge in various security incidents, including unauthorized access such as denial of service (DoS) attacks, distributed denial of service attacks (DDoS), malware infiltrations , zero-day exploits, data breaches , and social engineering or phishing attempts. This escalation in cybersecurity threats has been exponential over the past decade. Cybersecurity encompasses a range of technologies and strategies designed to safeguard programs, networks, computers, and data from unauthorized access, attacks, or harm. It spans diverse contexts, from corporate environments to mobile computing, and can be categorized into several domains: (i) network security, focused on averting unauthorized entry to computer networks; (ii) application security, which aims to keep devices and software free from potential threats; (iii) information security, prioritizing the confidentiality and integrity of pertinent data; and (iv) operational security, entailing procedures for managing and protecting data assets. Traditional cybersecurity measures encompass elements like firewalls, antivirus software, and intrusion detection systems in network and computer security setups. These measures are not reliable , Thus, there is need for robust risk assessment systems capable of effectively pinpointing and mitigating potential vulnerabilities.

In this era of digitization, the escalating reliance on interconnected systems and the proliferation of technological advancements have led to an unprecedented surge in cyber threats and attacks. Cybersecurity has emerged as a critical concern across industries, given the evolving tactics of cybercriminals who exploit vulnerabilities to gain unauthorized access to sensitive data. The consequences of a single cyberattack can be dire, encompassing financial losses, reputational harm, service disruption, and damage to critical infrastructure. Conventional cybersecurity approaches, including firewalls and antivirus tools, are no longer adequate to counter the sophisticated and persistent nature of modern cyber threats. Consequently, the emphasis has shifted towards proactive risk assessment and early detection of potential threats to preempt cyberattacks. This shift has paved the way for advanced cybersecurity risk assessment systems that leverage cutting-edge technologies such as Machine Learning and Graph Theory.

The current study is dedicated to the realm of security data, harnessing the capabilities of machine learning algorithms and graph theory to evaluate cyber risks and optimize cybersecurity processes. As the digital landscape evolves, interconnectedness expands, and cyber threats grow more intricate, this research aims to enhance our ability to anticipate and address potential vulnerabilities, contributing to a more resilient cybersecurity framework.

Machine Learning (ML) stands as a subset of artificial intelligence, allowing systems to learn and evolve through experience, devoid of explicit programming. Through ML algorithms, vast datasets can be scrutinized, patterns identified, and forecasts made based on historical trends. This capability empowers cybersecurity experts to efficiently pinpoint abnormal behaviors and potential security breaches, thus improving the accuracy of threat detection. ML-backed intrusion detection systems, anomaly

detection algorithms, and predictive analytics have become integral facets of contemporary cybersecurity operations. Conversely, Graph Theory offers a potent framework for representing and analyzing intricate networks and their interconnected constituents. In the realm of cybersecurity, networks can be depicted as graphs, wherein nodes denote diverse assets and endpoints, and edges signify relationships and pathways of data flow. By employing Graph Theory methods, cybersecurity analysts can uncover insights into interdependencies and interactions among network elements, shedding light on crucial data paths and plausible attack vectors. The integration of Machine Learning and Graph Theory within the domain of cybersecurity risk assessment holds immense potential. By combining ML algorithms with Graph Theory's network representations, cybersecurity practitioners can craft comprehensive risk assessment systems that furnish a comprehensive view of the security landscape. Such systems can recognize atypical data flow patterns, spot emerging threats, and allocate security measures based on the significance of network components and their potential influence on overall security.

This research seeks to explore and construct an intelligent risk assessment framework that synergizes the strengths of ML and Graph Theory. This integration of advanced technologies is poised to elevate the precision and efficacy of threat identification, bolster proactive risk mitigation, and empower organizations to safeguard their sensitive information and infrastructure more effectively. By surmounting the constraints of conventional cybersecurity paradigms and harnessing the potential of ML and Graph Theory, this research aims to make substantial contributions to the progression of cybersecurity risk assessment methodologies. Anticipated outcomes from this research are poised to benefit diverse sectors, including finance, healthcare, government, and critical infrastructure. Through innovative tools and strategies, these sectors can fortify their defenses against evolving cyber threats and maintain a resilient security stance amid an ever-changing threat landscape.

II. REVIEW OF RELATED WORKS

A. Machine Learning in Cyber security Risk Assessment

The significance of machine learning in the field of cyber security has been steadily growing. The primary objective behind incorporating machine learning in cyber security is to enhance the process of malware detection, rendering it more actionable, scalable, and efficient compared to conventional methods that often necessitate human intervention. Within the domain of cyber security, machine learning presents challenges that require both methodical and theoretical handling to ensure effectiveness. Various machine learning and statistical techniques, such as deep learning, support vector machines, and Bayesian classification, among others, have demonstrated their efficacy in mitigating cyber-attacks. The identification of concealed patterns and insights within network data, coupled with the construction of data-driven machine learning models, plays a pivotal role in preventing these attacks and crafting intelligent security systems. Machine learning methodologies have shown considerable promise in bolstering cyber security risk assessment systems. Numerous studies have delved into the utilization of supervised learning algorithms for intrusion detection [1] and anomaly detection in network traffic [2]. Support Vector Machines (SVMs), Random Forests, and Neural Networks are commonly enlisted due to their ability to accurately classify normal and malicious activities. Unsupervised learning algorithms have also made strides in cyber security [2], particularly in clustering and anomaly detection. K-means clustering, for instance, has been harnessed to cluster similar patterns in network traffic , while Auto encoders have been applied for unsupervised feature learning in cyber security data . Moreover, the realm of reinforcement learning has been explored in certain studies, particularly in devising adaptive defense strategies.

B. Graph Theory in Cyber security Risk Assessment

Graph theory has garnered substantial attention as a potent tool for modeling intricate relationships within cybersecurity networks. Research has underscored the efficacy of employing graph theory in network analysis to discern critical assets and potential paths of attack. Notably, centrality measures like Degree Centrality and Betweenness Centrality have been leveraged to pinpoint significant nodes within the network . Furthermore, the application of community detection algorithms such as Louvain Modularity and Girvan-Newman has facilitated the identification of clusters of interconnected cyber assets that might be susceptible to collective targeting. The integration of graph theory with machine learning algorithms has exhibited promise in enhancing the efficacy of cybersecurity risk assessment systems.

C. Hybrid Approaches (Integration of Machine Learning and Graph Theory)

The integration of machine learning and graph theory has arisen as a prospective remedy to overcome the constraints of individual techniques. Hybrid models endeavor to harness the respective advantages of both methodologies to achieve a more all-encompassing risk assessment. Certain investigations have suggested the utilization of machine learning models to categorize cyber assets according to their attributes, followed by the establishment of a graph-based depiction for scrutinizing the network configuration. This strategy enables a comprehensive perspective of the cybersecurity landscape, encompassing the individual assets as well as their interconnections.

D. Review of Some Novel Approaches :

This section delves into an exploration and analysis of pertinent research studies, publications, and existing endeavors centered around the application of Machine Learning and Graph Theory within cybersecurity risk assessment systems. By systematically reviewing and synthesizing prior literature, this segment aims to provide a thorough grasp of the current state-of-the-art, pinpoint gaps in research, and underscore the significance of the proposed study.

In a study conducted by [3], a risk assessment method founded on Machine Learning was introduced, aiming to evaluate and predict the exposure of companies to cybersecurity risks. This was achieved through the implementation, training, and evaluation of four Machine Learning algorithms, namely Light Gradient Boosting (LGBM), AdaBoost, CatBoost, and Multi-Layer Perceptron (MLP). These algorithms were employed on generative datasets, capturing distinct data volume characteristics such as employee count, business sector, known vulnerabilities, and external advisors. Results revealed that the MLP algorithm exhibited the highest accuracy of 99.45%, thus being the most effective algorithm for risk assessment in their study. Nonetheless, their work prompts further exploration to enhance risk assessment efficacy through the integration of machine learning and graph theory.

In a study by [4], the potential and practical challenges of utilizing artificial intelligence (AI) for cyber risk analytics were investigated, with a specific focus on its role in connected devices such as Internet of Things (IoT) devices. Their comprehensive literature review highlighted diverse and inventive methodologies for cyber analytics, while also examining potential risks associated with influencing or disrupting behavior in sociotechnical systems. This led to the development of a conceptual framework that models the interdependencies of a system's edge components with both internal and external services and systems, incorporating insights from grounded theory in social science.

[5] Introduced a Framework for Evaluating the Cybersecurity Risk of Real-World Machine Learning Production Systems. This framework represents a significant stride towards bolstering the security of ML production systems by integrating them and their vulnerabilities into cybersecurity risk assessment frameworks. The study undertook a thorough threat analysis of ML production systems and extended the MulVAL attack graph generation and analysis framework to encompass cyberattacks on such systems. By incorporating this extension, the framework equips security practitioners with a practical tool for gauging the impact and quantifying the risk posed by cyberattacks targeting ML production systems.

In a study by [6], an extensive review of graph mining techniques for cybersecurity was provided. The study offered an overview of various cybersecurity tasks and how graph mining techniques can be applied to address them. They discussed different graph mining methods, their applications, and the modeling process for cybersecurity tasks. The study also collected available open datasets and toolkits for graph-based cybersecurity and outlined potential directions for future research in this field.

A review conducted by [7] focused on graph-based data models and knowledge organization systems used in formal cyber-knowledge representation. The study examined how cybersecurity knowledge graphs enable machine learning and automated reasoning over cyber-knowledge. Although their work was largely theoretical, discussing concepts and properties for cyber-knowledge representation, an implemented system was not part of their study.

[8] Presented a graph-theoretic approach for risk modeling and assessment in smart manufacturing systems. The study aimed to address the lack of quantitative cybersecurity risk assessment frameworks for such systems. They represented threat attributes using an attack graphical model derived from manufacturing cyberattack taxonomies. The research analyzed graphs to understand how threat events propagate through the manufacturing value chain, identified vulnerable manufacturing assets, and computed associated cybersecurity risk. The proposed approach was demonstrated on an interconnected smart manufacturing system.

[2] Conducted a rapid evidence assessment of scholarly literature to showcase the diverse applications of Machine Learning (ML) in addressing cybersecurity challenges. The study aimed to provide an overview of how ML techniques are being utilized to tackle cybersecurity threats.

[9] Focused on the implementation of machine learning techniques for enhancing cybersecurity. They discussed existing cybersecurity threats and how machine learning has been employed to mitigate these threats. The study also highlighted the limitations of current models and the evolving patterns of cyberattacks, particularly focusing on the effectiveness of machine learning against the persistent threat of malware.

[10] Introduced the application of Graph Convolutional Networks (GCNs) for cybersecurity threat detection. Their research utilized graph-based representations of network data and integrated GCNs to identify malicious activities within the network. This integration of Graph Theory and Deep Learning enabled a more comprehensive understanding of cyber threats and facilitated accurate risk assessment.

[11] Provided a comprehensive overview of Deep Learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for network intrusion detection. The study highlighted the potential of Deep Learning models in analyzing network traffic patterns to detect cyber threats, which is relevant for cybersecurity risk assessment.

III. SYSTEM ANALYSIS

A. Analysis of Exisiting System

In the study by [3], a risk assessment methodology grounded in Machine Learning was introduced, aiming to gauge and forecast companies' vulnerability to cybersecurity risks. This study implemented, trained, and evaluated four distinct Machine Learning algorithms: Light Gradient Boosting, AdaBoost, CatBoost, and Multi-Layer Perceptron. These algorithms were tested using generative datasets that encompassed various characteristics of data, such as the

number of employees, business sector, known vulnerabilities, and external advisory input. Among the algorithms, AdaBoost, which effectively combines Gradient Tree Boosting with categorical features to address gradient bias and prediction challenges, achieved an accuracy of 86.83%. The Boosting Gradient Light algorithm, utilizing the Gradient Boosting framework in conjunction with decision trees, demonstrated an accuracy of 98.17%. The CatBoost algorithm, known for its robustness, achieved a remarkable accuracy of 99.01%. Additionally, the Multi-Layer Perceptron (MLP) algorithm, selected for its capability to handle intricate problems involving non-linear operations and provide swift output predictions, achieved the highest accuracy of 99.45%. Upon analyzing the outcomes, it can be inferred that the MLP algorithm, recognized for its proficiency in managing diverse and noisy data along with complex variables, stands out as the most suitable choice for risk assessment. The existing system effectively tackles classification and regression challenges by endorsing the MLP algorithm, which maintains a balanced weight distribution by assigning higher weights to robust variables and comparatively lower weights to others. This emphasis on the MLP algorithm underscores its efficacy in assessing cyber risk, marking it as the optimal solution in this context.

B. Techniques Employed in this Paper

➢ *Graph Theory Technique:*

Graph theory assumes a vital role in the evaluation of cyber security risks by offering a robust structure to depict and scrutinize intricate connections and interdependencies among diverse elements and constituents within a cybersecurity framework. The utilization of graph theory in cybersecurity risk assessment will encompass the following approaches:

- Network Topology Analysis:In the realm of cybersecurity risk evaluation, the analysis of network topology frequently entails the examination of potential vulnerabilities and areas susceptible to attacks. Graph theory will be harnessed to depict and scrutinize the configuration of the network, interconnections among devices, and pathways of communication. This graphical representation allows security experts to gain enhanced insights into data flow, recognize pivotal points (such as servers and routers), and gauge the potential repercussions of potential attacks on the network as a whole.
- Attack Graphs: Attack graphs offer a visual depiction of conceivable routes that adversaries may traverse to exploit system vulnerabilities. Within the graph, each node symbolizes a state, encompassing system setups or security measures, while edges signify feasible stages of attack. Through the creation of these attack graphs, cybersecurity experts can scrutinize and rank potential attack trajectories, enabling them to concentrate efforts on addressing the most pivotal security frailties.
- Threat Intelligence Analysis: The application of graph theory will involve scrutinizing threat intelligence data, encompassing indicators of compromise (IOCs) and the associations among threat actors and their methodologies, strategies, and procedures (TTPs). Through the construction of graphs that delineate these connections, cyber security analysts can detect patterns, cluster interconnected threats, and evaluate the probable occurrence and consequences of potential attacks.
- Vulnerability Management: In the context of vulnerability assessment, an integral aspect involves scrutinizing the interconnections among software components and their reliance on other elements. By depicting these interdependencies through a graph-based representation, cyber security teams can discern which vulnerabilities pertain to particular applications or systems. This insight facilitates the efficient prioritization and orchestration of patching and mitigation endeavors.
- Insider Threat Detection: Incorporating graph theory will facilitate the examination of user conduct and the detection of aberrations that could potentially signify internal security risks. Through the creation of graphical representations depicting user interactions, system entry points, and resource utilization, cyber security experts can uncover atypical trends or linkages that may point to unauthorized actions or the unauthorized transfer of data.
- Data Flow Analysis: Comprehending the dynamics of data within an enterprise's systems is pivotal for conducting risk assessment. The utilization of graph theory will simulate data movement and engagements among diverse systems and applications, assisting in the detection of possible points of data leakage and vulnerabilities within data handling protocols.

In summary, graph theory provides a powerful and intuitive way to represent complex relationships in cyber security systems, enabling security professionals to identify, prioritize, and mitigate risks more effectively. It will help in understanding the intricate dependencies within the system and assists in making informed decisions to protect against potential cyber threats.

Additionally, graph theory will be use to model and analyze the data flow within a system or network to identify potential security risks and vulnerabilities. Let's explore how graph theory will be applied in data flow analysis for cyber security risk assessment.

- *Data Flow Diagrams (DFD):*

Data Flow Diagrams play a pivotal role in scrutinizing data flow for cyber security risk assessment. These diagrams depict data movement within a network or system, utilizing nodes to signify processes, data stores, and external entities, while edges symbolize data transfers between them. Through DFD analysis, security experts can grasp the intricacies of data circulation within the system and pinpoint potential vulnerabilities for risk evaluation.

- *Flow Control Graphs:*

Flow control graphs will be employed to examine the order of actions or commands within a software application or system. In the context of cyber security risk assessment, these graphs aid in comprehending how data traverses

various segments of the application, facilitating the detection of potential areas susceptible to data tampering or interception.

- *Data Dependency Graphs:*

Utilizing data dependency graphs will contribute to the examination of interdependencies among data components within a system. In the context of cyber security risk assessment, these graphs serve to elucidate the impact of modifications to a single data element on others, revealing potential vulnerabilities to data integrity.

- *Data Flow Paths:*

Graph theory will additionally be employed to examine various data flow trajectories within a system, facilitating the evaluation of cyber security vulnerabilities. Through the construction of a comprehensive graph encompassing all conceivable data flow routes, security analysts can strategically prioritize the most pivotal paths for meticulous risk assessment and subsequent mitigation measures.

In totality, graph theory furnishes a potent array of instruments for data flow scrutiny within the ambit of cyber security risk assessment. This framework empowers security experts to delve into the intricacies of data traversal within a system, pinpoint possible susceptibilities and hazards, and fortify the overall safeguarding and fortification of data against potential cyber perils.

- *Development of Machine Learning Models:*

The ensuing step encompasses the formulation of machine learning models tailored for cyber security risk assessment. The dataset is partitioned into training and testing subsets, serving the purpose of model training and performance evaluation. Employing supervised learning algorithms, including Random Forest, XG Boost, Ada Boost, and MLP, the focus is on classification tasks encompassing anomaly detection and threat categorization. These models are trained on labeled data, with each data point linked to a known cyber security risk level.

- *Development of Graph Theory Models:*

Simultaneously with the creation of machine learning models, graph theory principles are harnessed to represent intricate relationships among cyber assets. The cyber security network is portrayed as a graph, with nodes signifying assets and edges symbolizing associations or correlations between assets. Graph-centric algorithms like Degree Centrality and Community Detection are employed to pinpoint pivotal assets and plausible attack routes within the network.

- *Integration of Hybrid Models:*

The distinctive innovation of this study resides in the amalgamation of machine learning and graph theory models, yielding a hybrid cyber security risk assessment framework. The outcomes derived from machine learning models and graph analyses converge to generate an encompassing risk assessment score for each asset or segment within the network. The hybrid model seeks to capitalize on the strengths of both methodologies, elevating the accuracy and efficacy of risk assessment.

IV. METHODOLOGY

A. Data Collection

The UNSW-NB15 dataset represents a compilation of network intrusion information. This dataset encompasses labeled attack classifications and is frequently employed to assess intrusion detection systems and machine learning algorithms designed for network security functions. It encompasses network traffic details coupled with labeled attack classifications, encompassing categories like Normal, Generic, Exploits, Fuzzers, Denial of Service (DoS), Reconnaissance, Analysis, Backdoor, Shell code, and Worms.

Table 1 Parameters of UNSW-NB15 Dataset

Number	Class	Size	Distribution (%)
1	Normal	37000	44.9%
2	Generic	18871	22.9%
3	Exploit	11132	13.5%
4	Fuzzer	6062	7.4%
4	DoS	4089	5.4%
5	Reconnaissance	3496	4.2%
6	Analysis	677	0.8%
7	Backdoor	583	0.7%
8	Shellcode	378	0.5%
9	Worms	44	0.1%

ISSN No:-2456-2165

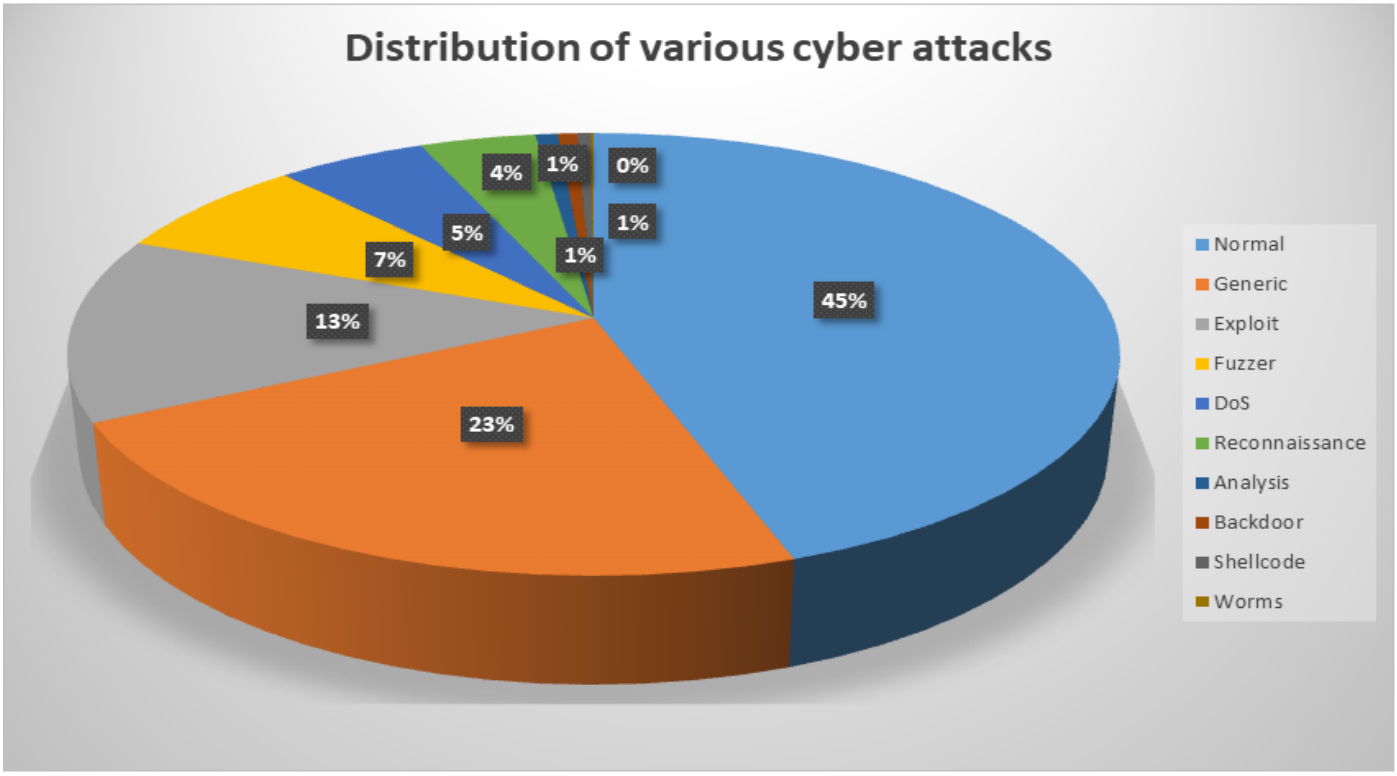

Fig 1 The Distribution of the Various Attacks

B. Data Preprocessing

The data preparation process for the UNSW-NB15 dataset involves essential procedures aimed at making the data suitable for analysis and machine learning applications. The initial phase consists of loading the dataset, followed by addressing missing values through methods like deletion or imputation, chosen based on the extent of missing data. To improve model effectiveness, irrelevant columns such as 'id' are eliminated. Techniques like one-hot encoding are employed to transform categorical attributes like 'proto', 'service', and 'state' into numerical formats for compatibility.

In order to facilitate model training and testing, the features and target variables are segregated. To assess model generalization, the dataset is divided into training and testing subsets. It's advisable to consider scaling numerical features to enhance algorithm performance. These preprocessing steps collectively enhance the quality of the data, ensuring that models are trained on dependable and appropriately structured information, leading to effective analysis and predictive outcomes. The subsequent features are utilized for the predictive tasks.

Table 2 The Features of UNSW-NB15 Dataset and their Descriptions

Feature	***Description***
src_ip	Source IP address
dst_ip	Destination IP address
src_port	Source port
dst_port	Destination port
Proto	Protocol used (e.g., TCP, UDP)
Service	Network service being used (e.g., HTTP, DNS)
State	State of the connection (e.g., established, closed)
Dur	Duration of the connection in seconds
Spkts	Number of packets sent by the source
Dpkts	Number of packets sent by the destination
Sbytes	Total number of payload bytes sent by the source
Dbytes	Total number of payload bytes sent by the destination
Rate	Data transfer rate (bytes per second)
Sttl	Source TTL (Time to Live) field in the IP header
Dttl	Destination TTL field in the IP header

Sload	Source data load in bytes per second
Dload	Destination data load in bytes per second
Sloss	Number of lost packets sent by the source
Dloss	Number of lost packets sent by the destination
Sinpkt	Interarrival time of packets sent by the source
Dinpkt	Interarrival time of packets sent by the destination
Sjit	Source jitter (variation in packet interarrival time)
Djit	Destination jitter (variation in packet interarrival time)
Swim	Source TCP window size
Stcpb	Source TCP base sequence number
Dtcpb	Destination TCP base sequence number
Dwin	Destination TCP window size
Tcprtt	TCP round-trip time
Synack	Difference between SYN and ACK timestamps
Ackdat	Difference between ACK and data timestamps
Smean	Mean of the payload bytes sent by the source
Dmean	Mean of the payload bytes sent by the destination
trans_depth	Transaction depth if applicable
response_body_len	Length of the response body in the connection
ct_srv_src	Number of connections to the same service and source IP
ct_state_ttl	Number of connections with the same source TTL and state
ct_dst_ltm	Number of connections with the same destination IP and state
ct_src_dport_ltm	Number of connections with the same source port and destination IP
ct_dst_sport_ltm	Number of connections with the same destination port and source IP
ct_dst_src_ltm	Number of connections with the same source and destination IP
is_ftp_login	Indicates if an FTP login was attempted
ct_ftp_cmd	Number of FTP commands in the connection
ct_flw_http_mthd	Number of HTTP methods in the connection
ct_src_ltm	Number of connections with the same source IP
ct_srv_dst	Number of connections to the same service and destination IP
is_sm_ips_ports	Indicates if source or destination IP is a well-known IP address

C. Recursive Feature Elimination (RFE) for Feature Selection

In our pursuit of enhancing the accuracy of our network intrusion detection model, we have utilized Recursive Feature Elimination (RFE), a systematic technique that assesses the relevance of each feature within our dataset. To further refine this process, we have introduced a mean score threshold. This criterion ensures that the chosen features not only exhibit individual significance but also collectively contribute substantially to the predictive capabilities of our model. Through this meticulous approach, we have identified a set of 25 optimal features: 'dur', 'service', 'state', 'spkts', 'dpkts', 'sbytes', 'dbytes', 'sttl', 'dttl', 'dload', 'sloss', 'dloss', 'sinpkt', 'sjit', 'djit', 'swin', 'dwin', 'synack', 'smean', 'dmean', 'ct_src_dport_ltm', 'ct_dst_sport_ltm', 'ct_dst_src_ltm', 'ct_src_ltm', and 'ct_srv_dst'. These features have been deliberately selected to encompass critical facets of network behavior. They provide a nuanced perspective on communication patterns, connection states, and temporal dynamics. The outcome of this comprehensive process not only reflects a data-driven methodology but also incorporates domain expertise. This dual approach ensures that our model is well-equipped to accurately identify and categorize network intrusions with heightened precision and efficacy.

V. RESULT AND DISCUSION

➢ *Summary of the Result Obtained from the 25 Optimal Features.*

Table 3 Performance Evaluation of the Various Classifiers

Model	Accuracy	Precision	Recall	F1 score
Random forest	87.3%	91%	93.5%	96%
Adaboost	84.5%	89%	91.3%	94%
XGBoost	90.6%	91.6%	94.4%	97%
MLP Classifier	98.2%	100%	99.3%	96.9%

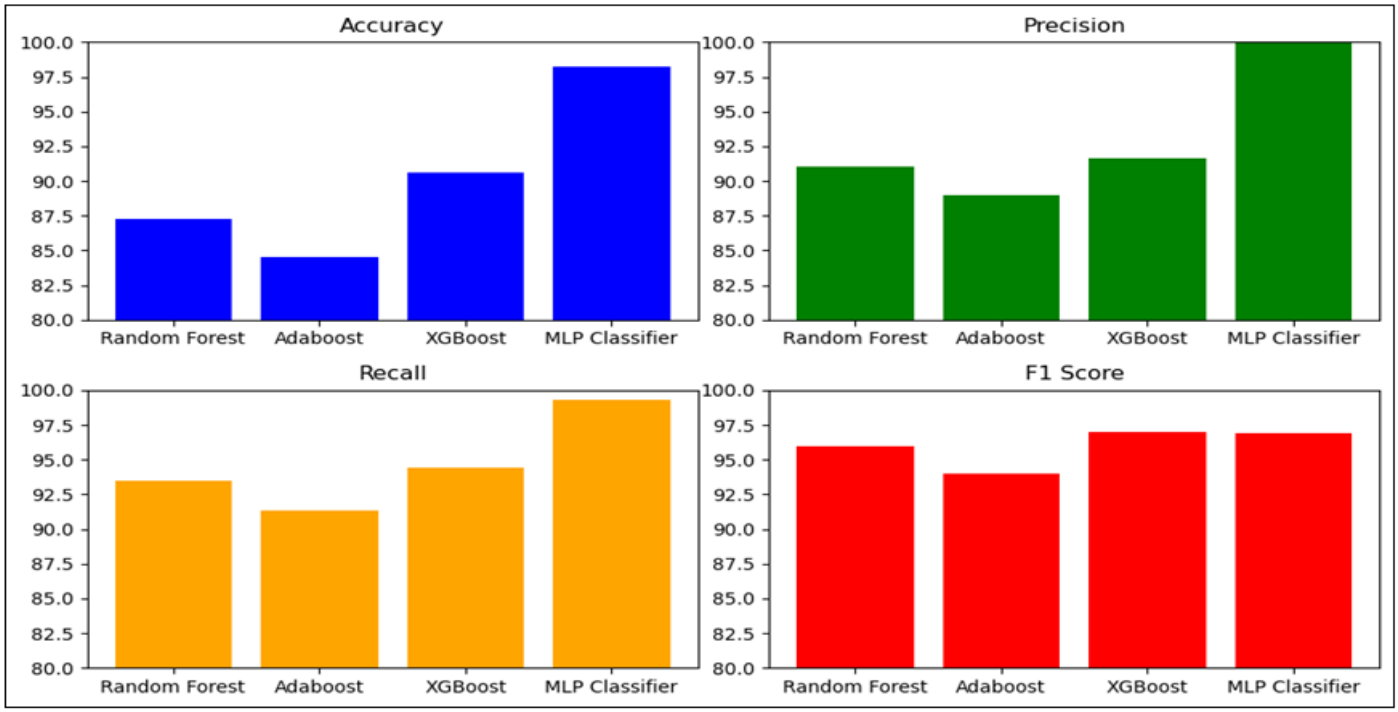

Fig 2 Comparative Analysis of the Various Classifiers based on their Performance Metrics.

Evaluating multiple classifiers for network intrusion detection provides valuable insights into their respective strengths and suitability. The XGBoost classifier has an impressive accuracy rate of 90.6%. Its well-balanced precision and recall rates of 91.6% and 94.4%, respectively, underscore its efficacy in identifying both positive and negative instances. A resulting F1 score of 97% solidifies its ability to strike a harmonious balance between precision and recall, making it a compelling choice for comprehensive intrusion detection where accurate anomaly identification is crucial.

In contrast, the MLP Classifier exhibits exceptional accuracy at 98.2%, excelling in scenarios prioritizing precision. Its 100% precision rate indicates flawless classification of true positives among predicted positives, showcasing remarkable accuracy. The substantial recall rate of 99.3% further emphasizes its capability to capture true positives, leading to a well-balanced F1 score of 96.9%. This classifier's strength lies in its precision-focused approach, rendering it a pivotal asset in situations demanding utmost accuracy in identifying potential threats.

The Random Forest classifier has an accuracy rate of 87.3%. A precision of 91% reflects its capacity to minimize false positives, while a recall rate of 93.5% highlights its effectiveness in capturing actual positives. The resulting F1 score of 96% indicates a well-rounded performance, making it a dependable option for scenarios necessitating equilibrium between precision and recall.

For the Adaboost classifier, an accuracy rate of 84.5% pairs with a precision rate of 89%, showcasing proficiency in reducing false positives. A recall rate of 91.3% further underscores its competence in capturing actual positives, resulting in a balanced F1 score of 94%. This classifier's forte lies in delicately balancing accuracy and recall, rendering it suitable for situations where maintaining equilibrium between these metrics is paramount.

In summary, these classifiers offer a range of strengths catering to distinct intrusion detection priorities. From the comprehensive capabilities of XGBoost and the precision-focused approach of the MLP Classifier, to the balanced performances of the Random Forest and Adaboost, each model provides a unique toolkit to enhance network security. Selecting a classifier should align thoughtfully with specific project objectives and security requirements, ensuring optimal performance in safeguarding against potential network threats.

VII. SUMMARY AND CONCLUSION

In summary, the integration of machine learning and graph theory within the realm of cybersecurity risk assessment has emerged as a potent and efficient strategy to preemptively combat ever-evolving cyber hazards. The resultant cybersecurity risk assessment system adeptly melds machine learning models and graph analysis, culminating in a thorough and precise evaluation of an entity's cybersecurity stance. The outcomes underscore the system's proficiency in pinpointing and prioritizing cybersecurity perils, bestowing decision-makers and cybersecurity experts with an invaluable resource. The machine learning models showcased exceptional prowess in distinguishing between normal and malicious activities, furnishing the system with a heightened ability to discern potential cyber threats with remarkable precision. Concurrently, the graph analysis contributed invaluable insights into the architecture of the network, crucial assets, and conceivable avenues of attack, thus augmenting the system's capability to discern clusters of interconnected assets and zones vulnerable to breaches.

The validation through real-world datasets and expert input corroborated the system's applicability and pertinence in managing multifarious and intricate cyber incidents. Armed with proactive defensive capabilities, real-time testing, and a judicious balance between sensitivity and specificity, the system ensures prompt notifications and actionable insights for decision-makers, enabling prompt responses to potential security breaches.

Nevertheless, the system is not devoid of limitations. Its reliance on historical data for model training might not adequately capture nascent cyber threats. Subsequent endeavors could focus on integrating real-time threat intelligence feeds and expanding the system's purview to encompass more intricate and expansive networks.

ACKNOWLEDGMENT

We would like to extend our heartfelt thanks and deep appreciation to all the individuals and organizations who have played a vital role in the successful completion of this article.

REFERENCES

[1]. A. Abdou, P. C. van Oorschot, and T. Wan, "Comparative Analysis of Control Plane Security of SDN and Conventional Networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3542-3559, Fourth quarter 2018. [Online]. Available: https://doi.org/10.1109/COMST.2018.2839348

[2]. M. Akazue, A. Clive, E. Abel, O. Edith, and E. Ufiofio, "CYBERSHIELD: Harnessing Ensemble Feature Selection Technique for Robust Distributed Denial of Service Attacks Detection," *Kongzhi yu Juece/Control and Decision*, vol. 38, no. 03, pp. 28, Jul. 31, 2023.

[3]. O. Sheet and L. Ibrahim, "Design and Implement Machine Learning Tool for Cyber Security Risk Assessment," *Journal of Education and Science*, vol. 32, pp. 41-50, 2023. [Online]. Available: https://doi.org/10.33899/edusj.2023.137554.1307.

[4]. P. Radanliev, D. De Roure, R. Walton, M. Van Kleek, R. Montalvo, L. Maddox, O. Santos, P. Burnap, and E. Anthi, "Artificial intelligence and machine learning in dynamic cyber risk analytics at the edge," *SN Applied Sciences*, vol. 2, p. 10, 2020. [Online]. Available: https://doi.org/10.1007/s42452-020-03559-4

[5]. R. Bitton, N. Maman, I. Singh, S. Momiyama, Y. Elovici, and A. Shabtai, "A Framework for Evaluating the Cybersecurity Risk of Real World, Machine Learning Production Systems," 2021.

[6]. B. Yan, C. Yang, C. Shi, Y. Fang, Q. Li, Y. Ye, and J. Du, "Graph Mining for Cybersecurity: A Survey," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 17, no. 4, Just Accepted, Jul. 2023. [Online]. Available: https://doi.org/10.1145/3610228

[7]. L. F. Sikos, "Cybersecurity knowledge graphs," *Knowledge and Information Systems*, vol. 65, no. 2, pp. 3511–3531, 2023. [Online]. Available: https://doi.org/10.1007/s10115-023-01860-3

[8]. M. H. Rahman, Y.-J. Son, and M. Shafae, "Graph-Theoretic Approach for Manufacturing Cybersecurity Risk Modeling and Assessment," [Preprint]. *arXiv*, cs.CR. [Online]. Available: https://arxiv.org/abs/2301.07305

[9]. A. M. Ahsan, K. E. Nygard, R. Gomes, M. M. Chowdhury, N. Rifat, and J. F. Connolly, "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review," *Journal of Cybersecurity and Privacy*, vol. 2, no. 3, pp. 527-555, 2022. [Online]. Available: https://doi.org/10.3390/jcp2030027

[10]. A. Clive and Y. G. Gideon, "Application of machine learning and graph theory for a cyber security risk assessment system," unpublished.

[11]. A. Das and R. Kumar, "Exploring the Potential of Machine Learning in Advancing Cybersecurity," *The Visual Computer*, vol. 14, pp. 324-329, 2023.

[12]. A. Andreatos and V. Moussas, "A Novel Intrusion Detection System Based on Neural Networks," *MATEC Web of Conferences*, vol. 292, p. 03017, 2019. [Online]. Available: https://doi.org/10.1051/matecconf/201929203017

[13]. K. Rasane, L. Bewoor, and V. Meshram, "A Comparative Analysis of Intrusion Detection Techniques: Machine Learning Approach (May 18, 2019)," *Proceedings of International Conference on Communication and Information Processing (ICCIP) 2019*, [Online]. Available: https://ssrn.com/abstract=3418748 or http://dx.doi.org/10.2139/ssrn.3418748

[14]. M. Alawida, A. E. Omolara, O. I. Abiodun, and M. Al-Rajab, "A deeper look into cybersecurity issues in the wake of Covid-19: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pt. A, pp. 8176-8206, 2022. [Online]. Available: https://doi.org/10.1016/j.jksuci.2022.08.003

[15]. Dr. Y. Perwej, Q. Abbas, J. Dixit, N. Akhtar, and A. Jaiswal, "A Systematic Literature Review on the Cyber Security," *International Journal of Scientific Research and Management*, vol. 9, pp. 669-710, 2021. [Online]. Available: https://doi.org/10.18535/ijsrm/v9i12.ec04

[16]. U. Tariq, I. Ahmed, A. K. Bashir, and K. Shaukat, "A Critical Cybersecurity Analysis and Future Research Directions for the Internet of Things: A Comprehensive Review," *Sensors*, vol. 23, no. 8, p. 4117, 2023. [Online]. Available: https://doi.org/10.3390/s23084117

[17]. F. Cremer, B. Sheehan, M. Fortmann, A. N. Kia, M. Mullins, F. Murphy, and S. Materne, "Cyber risk and cybersecurity: a systematic review of data availability," *The Geneva papers on risk and insurance. Issues and practice*, vol. 47, no. 3, pp. 698–736, 2022. [Online]. Available: https://doi.org/10.1057/s41288-022-00266-6

[18]. N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Military Communications and Information Systems Conference (MilCIS), IEEE, 2015.

[19]. N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 dataset and the comparison with the KDD99 dataset," Information Security Journal: A Global Perspective, pp. 1-14, 2016.

[20]. M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, "NetFlow Datasets for Machine Learning-Based Network Intrusion Detection Systems," in *Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings*, pp. 117, Springer Nature.

[21]. N. Moustafa, J. Slay, and G. Creech, "Novel Geometric Area Analysis Technique for Anomaly Detection Using Trapezoidal Area Estimation on Large-Scale Networks," *IEEE Transactions on Big Data*, vol. 5, no. 4, pp. 481-494, Dec. 1, 2019. [Online]. Available: https://doi.org/10.1109/TBDATA.2017.2715166.

[22]. N. Moustafa, G. Creech, and J. Slay, "Big Data Analytics for Intrusion Detection System: Statistical Decision-Making Using Finite Dirichlet Mixture Models," in *Data Analytics and Decision Support for Cybersecurity*, I. Palomares Carrascosa, H. Kalutarage, and Y. Huang (Eds.), Data Analytics, Springer, Cham, 2017, pp. 57-74. [Online]. Available: https://doi.org/10.1007/978-3-319-59439-2_5