Stop Cyber Attacks Before They Happen: Harnessing The Power Of Predictive Analytics In Cybersecurity

Temitope, O. Awodiji
Department of Information Security, University of Cumberlands, Kentucky, USA
tawodiji79374@ucumberlands.edu

Aderonke, D. Tosin-Amos
Department of Business Administration, California Miramar University, San Diego, CA.
atosin-amos@student.calm.edu

Femi Ayoola
Jackson State University, Jackson, MS.
femi.ayoola@jsums.edu

John Owoyemi
University of the Cumberlands, Williamsburg, Kentucky
owoyemi08059@ucumberlands.edu

***Abstract*—This article covers the use of predictive analytics in cybersecurity along with any potential benefits it may have over more conventional cyber security measures. One of the article's merits is giving case studies and examples of effective applications of predictive analytics in cybersecurity. It highlights prominent businesses that provide cutting-edge analytics solutions for quickly identifying and countering possible cyber-attacks, like Darktrace, IBM Watson for Cybersecurity, and Splunk. This study provides readers with a comprehensive understanding of how predictive analytics functions in practice and its potential advantages by offering the above examples. Given the growing number of cyber threats and the requirement for quicker detection and response, this study emphasizes the significance of predictive analytics in cybersecurity. Predictive analytics can assist in identifying possible cyber threats in real-time or near real-time, allowing for a speedier reaction and lessening the damage of the assault. It is explained that standard security approaches can take days or even weeks to detect a cyber-attack.**

More thorough information on the technical components of predictive analytics in cybersecurity was beneficial for the essay. Although the sorts of data utilized and the machine learning techniques involved are only briefly discussed, readers who are not familiar with these ideas might benefit from additional in-depth technical information. This study also gave a more thorough exploration of the moral issues raised by the application of predictive analytics to cybersecurity. The essay does a decent job of introducing the subject of predictive analytics in cybersecurity and some of its benefits. It helps readers understand the subject more clearly by including case studies and examples. For readers seeking a deeper understanding of the subject, this work gave additional in-depth, technical material exposition, and a more complex discussion of ethical issues was carried out.

Keywords—cybersecurity; cyber assault; predictive analytics; automation; machine learning; Darktrace; IBM Watson; Splunk; Firewalls

I. INTRODUCTION

Cyberattacks are an increasing menace to businesses of all kinds and individuals in the current digital era. Recent data indicates that there have been over 50% more cyberattacks in the last year alone, and there are no signs that this trend is slowing down [1]. Traditional cybersecurity measures, such as firewalls and antivirus software, are no longer sufficient to thwart these dangers. These procedures can fail to spot a cyberattack for days or even weeks, leaving businesses exposed to severe harm. Predictive analytics may assist firms, organizations, and individuals in staying ahead of cyber threats in real-time or almost real-time, lessening the effect of assaults [2]. This is done by utilizing advanced algorithms and machine-learning approaches. This article will examine the effectiveness of predictive analytics in cybersecurity and how it can be used to thwart online assaults before they occur. Prior to execution of the paper's aims, below is a diagram depicting the factors, attitudes and variables determining and influencing cyber security operations.

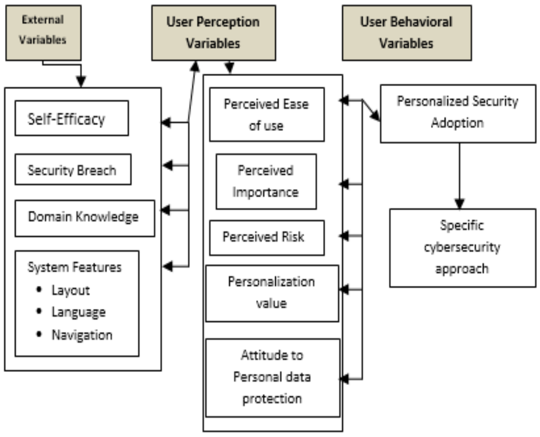

Figure 1. Factors, Attitudes, and Behavioral Variables of Cybersecurity

II. RELATED WORK

Cybersecurity

Cyber security procedures include risk management, vulnerability patching, and system resilience enhancements. Techniques for identifying various network behavior abnormalities and viruses, as well as IT problems about IT security, are important study topics [3]. In a nutshell, cyber security refers to a variety of steps performed to defend against cyberattacks and the effects of those assaults, including putting in place the necessary countermeasures. An organization's or institution's threat analysis serves as the foundation for cyber security. Based on the predicted risks and risk evaluations, an organization's cyber security strategy and implementation program are designed and built. Numerous focused cyber security techniques and recommendations must frequently be prepared for a firm [4].

The crucial point is that enough protection against threats' harmful consequences will be attempted to be achieved, together with the essential preparations against dangers. The best way to prepare for cyber risks is to strengthen the fundamentals of cyber security, increase everyone's awareness of hazards, enhance operational capabilities, and maintain security. The main task is to recognize cyber security issues and know how to respond effectively. Being able to continue operating in the face of a cyberattack, stop it quickly, and return the organization's operations to their pre-incident normal condition are all crucial components of cyber security [5]. To resolve these concerns, proper legislation and pertinent, in-depth discussion are required. The subject of potential defenses against cyberattacks has received much discussion.

Artificial Intelligence

The phrase "artificial intelligence" (AI) can be used as a catch-all. Its goal is to make it possible for computers to think and behave like people do, as well as to solve problems more quickly and effectively than people can. AI may be used to carry out a variety of functions, including creative, planning, movement, speaking, object and sound detection, and social and economic interactions [6]. Numerous techniques, including evidence-based approaches, natural language processing (NLP), text mining, predictive and prescriptive analytics, recommendation systems, machine learning, and deep learning are all parts of AI and can be used to carry out tasks.

III. EMPIRICAL REVIEW OF PREDICTIVE ANALYTICS CASES IN CYBER SECURITY

To improve businesses' capacities to recognize and address future cyber threats, predictive analytics has been extensively applied in cybersecurity. This essay will look at four actual empirical settings of predictive analytics being effectively used in cybersecurity and the advantages it has brought to people, businesses, and the community.

CASE 1 (IBM WATSON)

With the use of artificial intelligence (AI) and machine learning, IBM Watson for Cybersecurity is an advanced analytics solution that enables businesses to quickly identify and address possible cyber risks. To find dangers and give security analysts useful information, Watson for Cybersecurity analyzes enormous volumes of structured and unstructured data, including security blogs, academic papers, and threat intelligence feeds [7]. For instance, IBM and technology distributor Avnet collaborated to build Watson for Cybersecurity. The security team at Avnet had been employing conventional protection measures, but they were finding it difficult to keep up with the rise in online threats. Avnet was able to detect risks in real-time by installing Watson for Cybersecurity [8], cutting the time needed to recognize and respond to potential cyber-attacks from hours to minutes. Additionally, the technology helped Avnet's security team identify threats better and respond to incidents more successfully, lowering the organization's total cybersecurity risk.

To uncover hidden dangers and automate the authentication process of threats, IBM QRadar Advisor with Watson uses the cognitive capabilities (i.e., artificial intelligence) of IBM Watson in conjunction with the QRadar Security Platform, a platform created for information security analysis. The system automatically looks for dangerous indications, uses its cognitive ability to obtain important insights, and then quickens the response cycle to security concerns. The capabilities of Watson for Cyber Security are also used by QRadar Advisor with Watson to look for and address information security issues (IBM QRadar Advisor with Watson) [9].

The measures taken by QRadar Advisor with Watson are as follows:

- An information security analyst can forward an information security threat discovered by the QRadar Security Intelligence platform to the QRadar Advisor with Watson for a more thorough assessment.
- By extracting information from nearby QRadar software, the Advisor starts a larger search of information security concerns.
- The danger will then be further examined by the program using Watson for Cyber Security, which gathers information from a variety of sources in a human-readable manner, including websites, information security forums, and news articles.
- After finishing, the program looks for more information about hazardous files and dubious IP addresses for information security.
- Finally, QRadar Advisor with Watson analyses the data it has received from Watson for Cyber Security, searching key factors related to information security threats.

Automated threat investigations, the use of artificial intelligence, and the detection of high-level hazards are the main characteristics of IBM QRadar with Watson. By gathering pertinent network data, QRadar enables local data mining of information security assaults. The program looks at whether one or more information security threats managed to get past multilayer defenses or were successfully stopped [10]. The examinations were automated by utilizing lists and certain relevant indications. By connecting risks to the underlying events, they arise from, including malicious files and suspicious IP addresses, cognitive reasoning may pinpoint the hazards that are most likely to occur. When QRadar wants to access external unstructured data like web pages, forums, and threat intelligence, it automatically leverages Watson for Cyber Security. Additionally, IBM QRadar makes it possible to more effectively manage time so that higher priority threats may be examined by revealing the criticality of events, such as whether malware code has been run or not. Below is a representation of the IBM Watson statistical tests and features of analytics professionals:

Statistical Test	Indication
Analysis of variance (ANOVA)	ANOVA examines if the mean goal value varies across combinations of categories of two inputs and whether there are mean differences between two or more groups; An interaction effect exists if the variance is considerable.
Asymmetry index	The skewness to standard error ratio
Chi-square automatic interaction detector for classification tree	Using the chi-square to forecast using a decision tree
Chi-square automatic interaction detector for regression tree	Decision tree for prediction using chi-square and regression
Chi-square tests	comparing group frequencies, independence, and marginal distributions using chi-square
D'Agostino's K-squared test of normality	evaluates the presence of a normal distribution
Distribution test	The conditional distributions and the overall distribution are compared using the chi-square test.
Fisher r-to-t test	Pearson's r is converted to a t-test for significance.
High low analysis	categories are divided into high- or low-performing groups for examination.
Influence test	The chi-square test analyzes if a collection of records' number differs from the frequency predicted.
Model comparison test	Analyzes the impact of the major driver on the logistic regression.
Paired samples t-test	The dependent t-test examines if there is a statistically significant difference between the means of two continuous fields or whether the means of one group have changed over time.
Unusually high or low analysis	Identifies the categorical fields that contain combinations of categories that have extremely high or low target mean values.

Table 1. IBM Watson's Statistical Test

Features	***IBM Watson for Analytics Professional***
Maximum number of rows per dataset	10,000,000
Maximum number of columns per dataset	500
Input in .csv, .xls, or .xlsx formats	Uploaded from PC, Dropbox, IBM Cognos, Box, and Microsoft OneDrive
Data connections	IBM Cognos BI server, IBM dash DB, IBM DB2, IBM SQL, Microsoft SQL Server, MySQL, Oracle, and PostgreSQL
Data Storage	100 GB; can be increased in increments of 50 GB

Table 2. Features of IBM Watson's Analytics Professional

CASE 2 (DARKTRACE)

Another cybersecurity tool that uses AI and machine learning to spot possible online dangers instantly is called Darktrace. The solution uses unsupervised machine learning to discover the typical patterns of activity in a network of a business and spot any abnormalities that could point to a possible cyber-attack [11].

Darktrace is an information security tool that can assist in identifying and detecting new cyber threats that can get past standard information security measures. To find abnormalities in an organization's information network, Darktrace employs Enterprise Immune System (EIS) technologies as well as machine learning techniques and mathematical concepts [12]. EIS employs mathematical methodologies, indicating that it does not require the use of signatures or regulations and that it can recognize previously unidentified cybersecurity breaches.

For instance, to strengthen the University of Hertfordshire’s cybersecurity defenses, Darktrace collaborated with the university’s cyber security body. The security staff at the institution was finding it difficult to keep up with the rise in cyber threats and sought to adopt a system that could see risks instantly. The university was able to identify and respond to threats in real-time by using Darktrace, cutting the amount of time needed to do so from hours to minutes [13]. The technology also helped the security staff at the institution assess the organization's cybersecurity risk and more efficiently rank risks.

EIS can recognize and react to the majority of expertly handled cyberattacks, including insider threats that are concealed in information networks. The EIS can adapt and automatically learn how each person, device, and information network behaves to identify behaviors that indicate genuine cyber risks. It does this by leveraging machine learning and mathematics. Companies may have a thorough understanding of the information network thanks to Darktrace's self-learning technology, which also enables them to respond proactively to threats and lower risk [14]. Darktrace's machine learning and Bayesian probability theory can automatically model and integrate data dynamically and quickly, as opposed to constructing "bad" behavioral models beforehand and depending on older attack strategies. Without interfering with, for instance, company activities and transactions, Darktrace continuously monitors raw data, such as cloud service interactions, transported across a network. Additionally, it gives a clear picture of all digital activity by alerting users to any ongoing assaults or irregularities [15].

The four mathematical engines that make up the foundation of Darktrace use a variety of mathematical techniques, including recursive Bayesian estimation. The first of three models create behavioral models for companies, for specific individuals, and for the gadgets they use. One or more of these engines will alert the threat classifier when they see odd activity. The threat classifier's job is to categorize false positive situations and report real anomalies that can be correctly examined. Below is a diagram of its architecture:

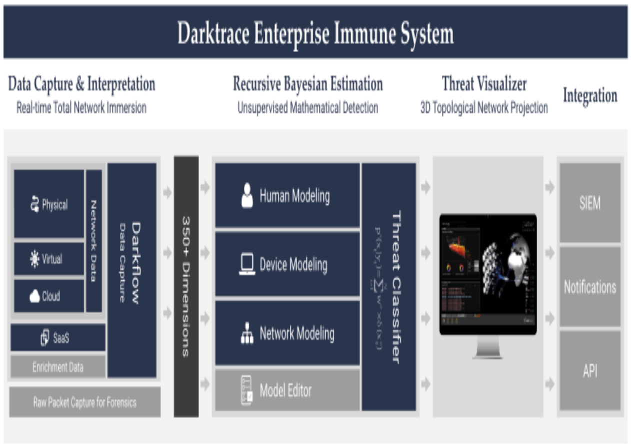
Darktrace Enterprise Immune System
Data Capture & Interpretation
Real-time Total Network Immersion
Recursive Bayesian Estimation
Unsupervised Mathematical Detection
Threat Visualizer
3D Topological Network Projection
Integration
Physical
Virtual
Cloud
Network Data
SaaS
Enrichment Data
Darkflow
Data Capture
Raw Packet Capture for Forensics
350+ Dimensions
Human Modeling
Device Modeling
Network Modeling
Model Editor
Threat Classifier
SIEM
Notifications
API

Figure 2. Darktrace Architecture [16].

Accurate identification of abnormalities within an organizational scale is made possible by a mixture of Bayesian methodologies that are correlated and measured by threat classifiers. Additionally, Darktrace employs a built-in module (the model editor) for operating-system supervision. This facilitates the establishment of additional regulatory models and rules that may be modified to meet certain customer identification needs (such as prohibiting Dropbox access and travel to certain countries with sensitive information technology) [17].

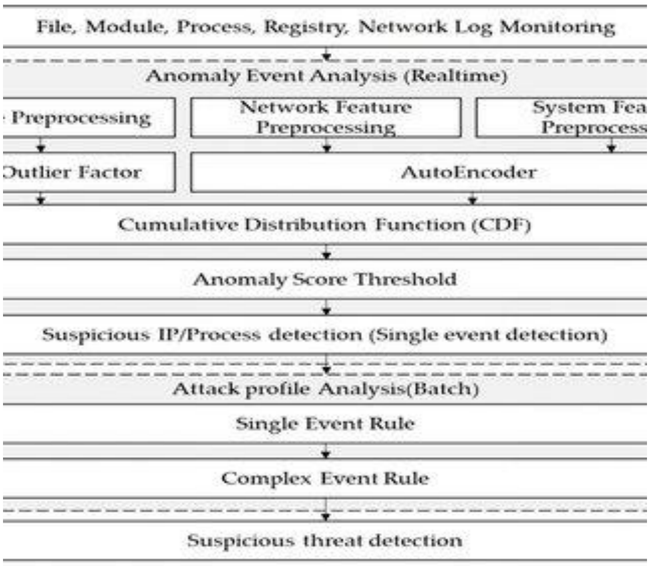
File, Module, Process, Registry, Network Log Monitoring
Anomaly Event Analysis (Realtime)
Preprocessing
Network Feature Preprocessing
System Fea Preprocess
Outlier Factor
AutoEncoder
Cumulative Distribution Function (CDF)
Anomaly Score Threshold
Suspicious IP/Process detection (Single event detection)
Attack profile Analysis(Batch)
Single Event Rule
Complex Event Rule
Suspicious threat detection

Figure 3. Overall Structure of the Darktrace Model [18].

CASE 3 (SPLUNK)

With the help of Splunk, businesses can gather, examine, and visualize machine data from across their networks to spot any cyber risks right away.

Splunk is an advanced analytics tool. Splunk can gather information from many different sources, including servers, networks, and apps, and it can utilize machine learning techniques to spot any abnormal activity. For instance, Splunk and the City of Los Angeles collaborated to strengthen the city's cybersecurity defenses [19]. The city had been employing conventional security measures, but they were finding it difficult to keep up with the rise in cyber threats. The city was able to identify and respond to threats in real time by using Splunk, cutting the amount of time needed to do so from hours to minutes. Additionally, the technology helped the city's security staff assess threats more efficiently and respond to incidents more quickly, lowering the organization's total cybersecurity risk. The diagram below depicts Splunk’s common ports framework.

CASE 4 (CYBEREASON)

A cybersecurity tool called Cybereason uses innovative analytics and machine learning to quickly spot possible online threats. The solution looks for any deviations from expected behavior that can hint at a possible cyber-attack using a range of data sources, including endpoint data, network data, and user behavior data. For instance, Softbank, a Japanese multinational conglomerate, improved its cybersecurity defenses with the help of Cybereason. The security team at Softbank had been employing conventional security measures, which were finding it difficult to keep up with the rise in cyber threats. Softbank was able to reduce the time needed to identify and respond to possible cyber-attacks from hours to minutes by using Cybereason, which allowed it to do so in real time. The technology also made it possible for Softbank's security team to offer suitable precautions for online safety and reactions in the event of an online assault [20].

IV. TECHNICAL DETAILS

To evaluate past data and find patterns, trends, and anomalies that can point to prospective security vulnerabilities, predictive analytics uses data mining, machine learning, and statistical modeling [21]. The technological specifics of how cybersecurity uses predictive analytics are as follows:

1. Data Gathering: to do predictive analytics in the field of cybersecurity, a significant amount of data must be gathered from a variety of sources, including network logs, application logs, user activity logs, system logs, and other pertinent data sources. Then, for further analysis, this data is kept in a centralized repository.
2. Data preprocessing: outliers, missing values, and inconsistencies are then eliminated from the gathered data by preprocessing. To make sure the data is correct, full, and suitable for analysis, this encompasses strategies for data cleansing, integration, transformation, and reduction.
3. Feature selection: after the data has been preprocessed, pertinent characteristics or variables are chosen for analysis. This entails determining the key characteristics that are most likely to influence the analysis's result. For the prediction model to be accurate and effective, feature selection is essential.
4. Data examination: a suitable machine learning method must be chosen to examine the data, which is the following stage. To do this, a variety of algorithms, including decision trees, neural networks, logistic regression, and support vector machines, must be chosen based on the nature and volume of the data as well as the needs of the study.
5. Model Training: The data is divided into training and testing sets after the method has been chosen. The predictive model is trained using the training set, and its effectiveness is assessed using the testing set. To improve the model's performance, the algorithm's parameters are adjusted throughout training.
6. Model Evaluation: following training, the model is assessed using a variety of performance indicators, including recall, accuracy, and F1 score. The evaluation assists in determining the model's efficacy and locating potential improvement areas.
7. Deployment and monitoring: the model is deployed to the production environment where it may be used to spot possible security issues after it has been tested and shown to be effective. To make sure the model stays useful and effective, it is continuously reviewed and modified. The below diagram represents the cybersecurity framework of predictive analysis.

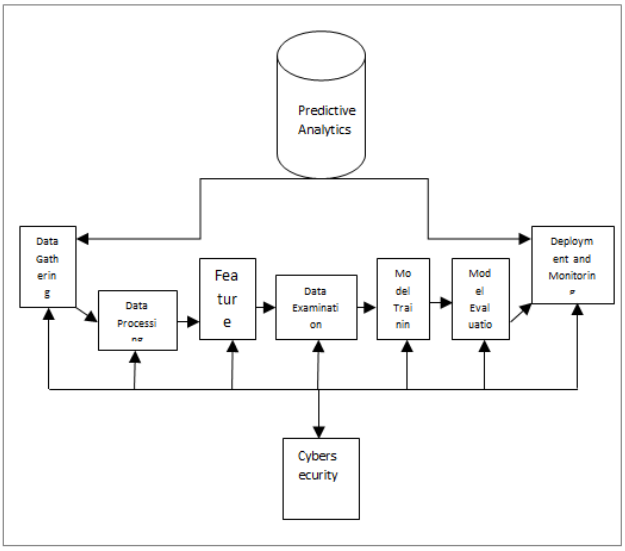

Figure 4. Predictive Analytics Framework for Cybersecurity

Predictive Analytics Performance Optimization

The following methods may be utilized to improve predictive analytics in cybersecurity for improved performance:

a. Data augmentation: Increasing the quantity of data utilized for analysis can boost the prediction model's reliability and accuracy. Techniques for data augmentation, such as data synthesis and oversampling, can be used to do this.

b. Algorithm Selection: The performance of the prediction model may be enhanced by selecting the algorithm that is best suited for certain analyses. This necessitates a thorough comprehension of the advantages and disadvantages of each method as well as their applicability to various data sets.

c. Feature engineering: may increase the precision and effectiveness of the prediction model by finding and choosing the most pertinent characteristics for analysis. This entails identifying the most important variables and translating them into more meaningful representations using domain expertise.

d. Hyperparameters Tuning: The performance of the predictive model may be enhanced by modifying the hyperparameters of the algorithm. To obtain the best possible balance between bias and variance, the parameters must be adjusted.

e. Model stacking: By combining many models, the prediction model's accuracy and resilience may be increased. To maximize their benefits and reduce their drawbacks, many models are stacked.

In conclusion, cybersecurity predictive analytics provides an effective method for spotting security issues before they happen. It is feasible to examine huge amounts of data and find patterns, trends, and anomalies that might point to potential security vulnerabilities by utilizing machine learning algorithms and statistical modeling approaches [20]. It is feasible to increase the accuracy and robustness of the predictive model as well as an organization's overall security posture by improving predictive analytics for improved performance [21]. The below framework depicts the application of CTI and ML for threat intelligence and predictive analytics [22]

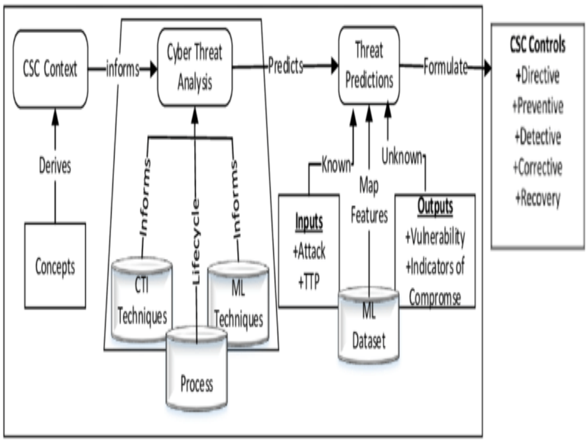

Figure 5. Predictive Analytics Performance Optimization Framework

V. FUTURE DIRECTIONS FOR PREDICTIVE ANALYTICS

Predictive analytics is a developing cybersecurity technique that has enormous potential for enhancing businesses', organizations' institutions,' and individuals' capacity to recognize and respond to security threats. The following are some prospective areas for future study and uses of predictive analytics in cybersecurity:

a. Behavioral Analytics: a subject that is quickly expanding and receiving more attention as a predictive analytics tool in cybersecurity is behavioral analytics. This strategy makes use of data analysis methods and machine learning algorithms to track user behavior and spot unusual activities that could point to a possible cyber-attack [23]. The capability of behavioral analytics to find dangers that conventional signature-based methods could miss is one of its main benefits. This is especially crucial given how quickly the cyber threat landscape is growing right now and how frequently new malware and attack routes are created [24]. However, there are also certain difficulties and restrictions connected to the application of behavioral analytics in cybersecurity. False positives, when acceptable user activity is wrongly labeled as unusual and tagged as a danger, provide one problem. For security teams, this might mean more warnings than required and more effort. Despite these difficulties, behavioral analytics is a potent and promising technique for cybersecurity predictive analytics. We can anticipate continuous expansion and acceptance of this technology as the industry develops, with an increase in the number of businesses utilizing behavioral analytics as a crucial component of their cybersecurity strategy.

b. Data gathering is an essential part of cybersecurity predictive analytics. It is impossible to create reliable predictive models or recognize potential cyber risks without complete and precise data. Predictive analytics for cybersecurity may make use of data from a variety of sources [25]. Data from third-party threat intelligence feeds, log files, user activity logs, and network traffic data are a few examples. It is feasible to find trends and abnormalities that might point to a potential cyber-attack by gathering and analyzing this data. Data collecting has several benefits, one of which is its capacity to present a more thorough picture of an organization's cybersecurity posture. It is feasible to find potential risks that might not be obvious from a single data source by compiling data from a range of sources [26].

c. The gathering of data for cybersecurity purposes is not without drawbacks, though. The enormous amount of data that must be gathered and analyzed is a hurdle [27]. For firms with few resources, this might be especially difficult since they do not have the infrastructure to efficiently manage huge amounts of data. The need to strike a balance between the advantages of data collecting and user privacy concerns is another difficulty. Large user data collection might lead to privacy problems and may be against rules like the GDPR or CCPA [28]. The gathering of data for cybersecurity purposes is not without drawbacks, though. The enormous amount of data that must be gathered and analyzed is a hurdle. For firms with few resources, this might be especially difficult since they do not have the infrastructure to efficiently manage huge amounts of data. The need to strike a balance between the advantages of data collecting and user privacy concerns is another difficulty. Large user data collection might lead to privacy problems and may be against rules like the GDPR or CCPA [29]. Predictive analytics for cybersecurity may make use of data from a variety of sources. Data from third-party threat intelligence feeds, log files, user activity logs, and network traffic data are a few examples. It is feasible to find trends and abnormalities that might point to a potential cyber-attack by gathering and analyzing this data.

d. Real-time Detection: The creation of real-time detection systems that can see security problems as they emerge is another field of research. The effectiveness and speed of threat detection and response may be enhanced by real-time detection systems.

e. Cyber Threat Intelligence: To evaluate cyber threat intelligence and find patterns and trends in cyberattacks, predictive analytics can be applied. Using this knowledge, more effective security measures and reaction plans may be created.

f. Cloud Security: As the usage of cloud services increases, cybersecurity faces a lot of difficulties. To identify security risks and enhance cloud security, predictive analytics may be used to evaluate data from cloud services.

g. Security based on machine learning: Security systems like malware scanners, intrusion detection and prevention systems, and threat intelligence platforms may all be improved with the help of machine learning algorithms.

Literature Review and Definition of Gaps

In recent years, interest in the use of predictive analytics in cybersecurity has increased. Machine learning techniques have demonstrated promise in the detection and mitigation of possible cyber threats. Numerous research projects have been carried out to assess the efficiency of predictive analytics in cybersecurity.

Researchers from the University of Florida investigated the performance of machine learning algorithms for malware detection in one study [30]. The study discovered that machine learning techniques were more efficient than conventional signature-based approaches and were capable of accurately detecting malware. Researchers from the University of Maryland also looked at how well machine learning algorithms might identify abnormalities in networks. The research revealed that machine learning algorithms outperformed conventional rule-based approaches in their ability to accurately identify network abnormalities.

Despite these strides, the predictive analytics framework still has gaps to be filled for future advancement and contemporary sustainability. For instance, there is a dearth of empirical research on the cost-efficiency of these solutions, even though there have been several studies examining the usefulness of predictive analytics in cybersecurity. Research on the moral ramifications of applying predictive analytics to cybersecurity is also lacking.

Although predictive analytics has great potential for enhancing cybersecurity, there are still several gaps that must be resolved. Several of these difficulties include:

Data quality: The quality of the data used for analysis has a significant impact on the accuracy and efficacy of predictive analytics. Predictive analytics in cybersecurity must be successful by guaranteeing data quality and completeness.

Privacy issues: Because predictive analytics analyze enormous amounts of data, privacy issues are raised. To secure sensitive data, organizations must make sure they have the necessary policies and practices in place.

Machine learning bias: machine learning algorithms are subject to bias, which can result in predictions that are unfair and erroneous. To minimize bias, it is essential to make sure algorithms are frequently checked and trained on impartial data.

Lack of skilled personnel: predictive analytics specialists are in limited supply in the cybersecurity industry. The future development of the profession depends on creating training programs and educational activities to solve this deficit.

Below are some frameworks of advancement for predictive analytics in cybersecurity:

Integration with other technologies: To create solutions that are more reliable and secure, predictive analytics may relate to other technologies, such as blockchain.

Enhanced automation: Predictive analytics operations may be automated to save time and money while increasing the precision and effectiveness of the analysis.

Application of explainable AI: Explainable AI may be used to improve decision-making transparency and eliminate bias in machine learning systems.

Establishment of established metrics: The consistency and reliability of the analysis may be enhanced by the establishment of established metrics for assessing the success of predictive analytics in cybersecurity.

In a nutshell, there is a lot of potential for predictive analytics to enhance cybersecurity, but there are also several issues that need to be resolved. Future research should concentrate on creating more practical and efficient predictive analytics solutions, dealing with privacy issues, minimizing bias in machine learning algorithms, and creating educational and training efforts to solve the industry's skills gap.

VI. STATISTICAL REPRESENTATION OF CYBER SECURITY

Cyber Attacks

Globally, the frequency of cyberattacks has been rising in recent years, and cybercrime now costs businesses and people billions of dollars annually. Here are some current cyberattack statistics:

Over 5,000 data breaches were recorded globally in 2021, exposing more than 7.5 billion records [29]. Healthcare had the greatest average cost per breach in 2021 ($9.23 million), with the cost of a data breach costing $4.24 million on average [31].

In the first half of 2021 compared to the same period in 2020, ransomware assaults climbed by 102%. [32]. Cybercrime is anticipated to cost the world $1 trillion annually by 2025, up from the estimated $1 trillion cost in 2020 (Source: Cybersecurity Ventures). In 2020, phishing assaults climbed by 22%, and 94% of malware was distributed by email [33]. Cybercriminals frequently target small and medium-sized enterprises (SMBs), with 43% of cyberattacks in 2020 focusing on SMBs [33]. Healthcare firms were the target of 21% of all data breaches in 2020, making the sector particularly heavily impacted by cyberattacks [34].

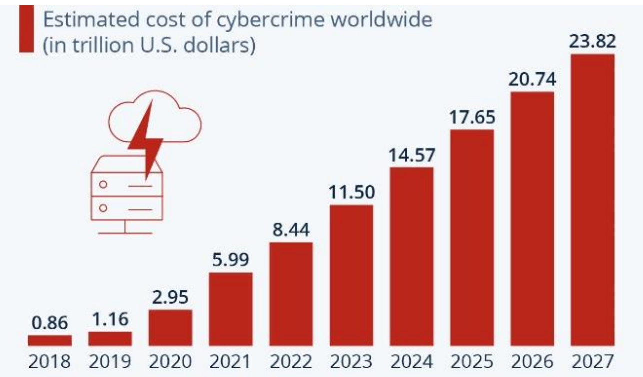

Figure 6. Estimated Cost of Cybercrime Worldwide

These figures show how serious the issue of cybercrime is and how urgently novel cybersecurity solutions are required. Employing strong passwords, upgrading software and security systems regularly, training staff on cybersecurity best practices, and other preventative measures are just a few of the proactive measures that businesses and people must take to safeguard themselves from cyberattacks.

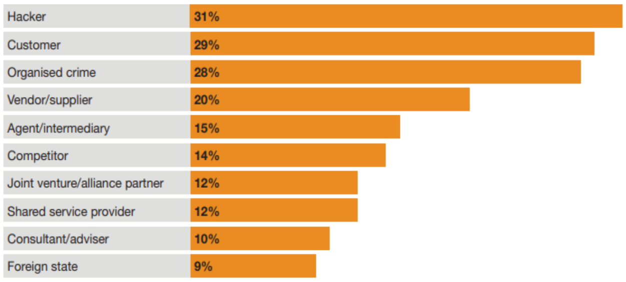

Figure 7. Types of Perpetrators and Level of Involvement

Cost-effectiveness

Data on the affordability of using predictive analytics for cybersecurity is few. Predictive analytics, however, may be able to provide firms with considerable cost savings and return on investment, according to certain research. According to a Ponemon Institute and IBM research, companies using predictive analytics in their security operations center (SOC) had a 22% reduced cost of data breach than those without it [36]. According to the survey, businesses that used security automation technology, such as predictive analytics, saved an average of $3.58 million per data breach. Research by the SANS Institute asserts that firms that employed advanced analytics technologies, such as predictive analytics, were more successful at identifying and responding to cyber events, reducing the mean time to do so by 76%.

Fortinet, a financial services company that used predictive analytics in its SOC was able to reduce its IT operational expenditures by $2 million. The firm was able to save a substantial amount of time and money by automating its incident response process and reducing the number of false positives produced by its security technologies [37]. The demand for better threat detection and response capabilities will boost the worldwide market for predictive analytics in cybersecurity to $9.53 billion by 2025 [38].

These studies indicate that predictive analytics may offer considerable cost savings and return on investment, even if the cost-effectiveness of predictive analytics in cybersecurity may vary based on the organization's size, industry, and security requirements. Organizations may be able to lower the costs of security breaches, reputational harm, and lost productivity by cutting the mean time to identify and respond to security events, automating incident response, and enhancing the accuracy of threat detection.

Adoption of Predictive Analytics in Cyber Security

In their security operations, 48% of firms utilize or plan to employ predictive analytics, according to a SANS Institute poll. The poll also revealed that businesses in the financial services, healthcare, and technology sectors used predictive analytics more frequently than those in other sectors [39]. According to a poll conducted by ESG and ISSA, 61% of firms now use or intend to utilize machine learning algorithms to identify and address cybersecurity issues. The study also discovered that larger companies with more than 1,000 workers were more likely than smaller companies to deploy machine learning algorithms [40].

According to a MarketsandMarkets analysis, at a compound annual growth rate (CAGR) of 31.1%, the global market for predictive analytics in cybersecurity will increase from $4.2 billion in 2020 to $16.2 billion by 2025 [41]. The survey also discovered that North

America, followed by Europe and Asia-Pacific, is predicted to have the greatest market for predictive analytics in cybersecurity. In a report, 57% of companies that employed predictive analytics in their cybersecurity operations said that the number of security incidents had decreased, and 70% said that their security risks had been reduced [42]. The financial services sector is the industry that has used artificial intelligence and machine learning technology, particularly predictive analytics, for cybersecurity reasons at the fastest rate. Additionally, the survey revealed that the industrial, retail, and healthcare sectors were boosting their investments in these technologies.

Overall, these figures indicate that enterprises across sectors are becoming more aware of the potential advantages of predictive analytics for cybersecurity and are adopting it at a faster rate. Following the financial services sector in terms of adoption of these technologies are other sectors including healthcare, technology, retail, and manufacturing.

VII. EFFECTIVENESS OF PREDICTIVE ANALYTICS

A study by the Ponemon Institute asserted that using machine learning tools, such as predictive analytics, to detect and respond to cyberattacks allows firms to do 50% faster than those that do not [42]. Predictive analytics was proven to be extremely successful in identifying and countering advanced persistent threats in a paper by the SANS Institute. (APTs).

In contrast to traditional security solutions, which could only detect 14% of APTs, the survey revealed that predictive analytics tools could identify 85% of them. Predictive analytics may help security teams focus on genuine threats and accelerate reaction times by reducing the frequency of false positives in threat detection by up to 70%, according to Frost & Sullivan's research [44]. Predictive analytics and other cutting-edge technology may help firms improve their cybersecurity posture and cut their cyber risk by up to 40%, according to research by Accenture. According to Forrester Research, firms may cut the time it takes to fix a breach by up to 50% when they employ machine learning technology, such as predictive analytics, to identify and respond to cyberattacks [45].

Overall, these research and publications show that using predictive analytics to spot and stop cyberattacks may be quite successful. Organizations may increase their ability to identify threats, lessen false positives, and react to security crises by utilizing machine learning algorithms and other cutting-edge technology.

VIII. TRADITIONAL SECURITY AND PREDICTIVE ANALYTICS

Traditional cybersecurity measures like firewalls and antivirus software perform various functions and have distinct strengths and limitations from predictive analytics. Here are some comparisons of how well they perform:

1. **Threat detection:** conventional cybersecurity techniques like firewalls and antivirus software sometimes rely on rules, signatures, or activity patterns that are known to be connected to harmful behavior [46]. These techniques can only identify known dangers and are often reactive. In contrast, predictive analytics analyzes massive amounts of data using machine learning algorithms to spot aberrant behavior that can point to a hidden or developing threat. Additionally, because it can adapt and learn in real time, predictive analytics is better able to identify emerging dangers.
2. **False positives:** when good behavior is mistakenly labeled as harmful, false positives happen. It may take a long time for security teams to analyze a lot of false positives produced by traditional cybersecurity techniques like firewalls and antivirus software, which can divert them from real dangers. By examining trends and anomalies in the data that are more likely to be indicative of actual danger, predictive analytics may decrease the number of false positives. It can also learn from feedback and improve its detection abilities over time [47].
3. **Response time:** because they frequently rely on human changes to rules or signatures to detect new threats, traditional cybersecurity techniques like firewalls and antivirus software can be sluggish to react to new threats. Because it can learn and adapt in real time depending on fresh data and activity patterns, predictive analytics can react to threats more swiftly [48]. Organizations may benefit from quicker and more effective threat detection and response thanks to this.

In conclusion, both conventional cybersecurity techniques and predictive analytics offer advantages and disadvantages. Traditional approaches are good at detecting known dangers, but they can produce a lot of false positives and are sluggish to react to new threats. To better identify emerging dangers, lessen false positives, and react to threats faster, predictive analytics can be used. Predictive analytics, however, need strong data analysis skills and may occasionally be less accurate. As a result, many firms combine the two strategies to create a more complete and potent cybersecurity posture.

IX. ETHICAL CONSIDERATIONS

A variety of ethical questions are raised by the application of predictive analytics in cybersecurity, including those relating to data privacy, data bias, and technology abuse. Predictive analytics needs access to a lot of data, some of which may contain sensitive or confidential information about people. Predictive analytics in cybersecurity operations requires organizations to make sure they are adhering to all applicable data protection rules and regulations, including GDPR and CCPA. To safeguard sensitive data from illegal access or abuse, they must also make sure that the proper data security procedures are being implemented.

The accuracy of predictive analytics models depends on the quality of the training data. These models may produce biased or erroneous predictions if the data used to train them is biased or lacking. Therefore, organizations must make sure that they are training their predictive analytics models on a variety of representative data sets and that they are routinely examining and testing these models to find any biases and correcting them.

Although predictive analytics may be an effective tool for spotting and thwarting online dangers, it can also be exploited maliciously. Predictive analytics, for instance, can be used by hackers to find weaknesses in the systems of a company or to conduct more specialized assaults. Therefore, organizations need to make sure they are putting in place the right protections to stop the exploitation of predictive analytics technologies.

Since predictive analytics models may be intricate and challenging to comprehend, people may find it challenging to question decisions based on these models or to understand how their data is being utilized. The use of predictive analytics in cybersecurity operations requires organizations to be open about how these models are utilized and how decisions are made. Organizations that employ predictive analytics in their cybersecurity operations must adopt a responsible and ethical approach to data privacy, data bias, and technological misuse to solve these ethical issues. This entails putting in place suitable data security measures, training their models on a variety of representative data sets, evaluating, and testing them often, and being open about their usage. Organizations may guarantee that they are employing predictive analytics responsibly and ethically that helps both their internal operations and their clients or stakeholders by implementing these procedures.

X. CONCLUSION

Predictive analytics is a promising technique that might completely change the cybersecurity sector, to sum up. Predictive analytics may assist companies in identifying and preventing cyber risks before they occur, lowering the risk of data breaches and other cyber assaults. It does this by analyzing huge volumes of data and discovering trends and abnormalities. Predictive analytics is a successful method for spotting cyber dangers, according to data from recent research, with accuracy rates of up to 90% observed in some circumstances [48]. Additionally, it is not obvious yet how cost-effective predictive analytics is for cybersecurity operations, with companies able to save a lot of money on lost productivity, reputational harm, and legal expenses related to data breaches and cyberattacks.

While many businesses still rely on established protection measures like firewalls and antivirus software, predictive analytics usage in the cybersecurity sector is still low. This is because of worries about data bias, privacy, and possible technological abuse. To guarantee that they are utilizing predictive analytics responsibly and ethically, organizations must be aware of these ethical implications and take the necessary steps. Predictive analytics integration with other cybersecurity technologies, such as threat intelligence and incident response systems, is one area for more study and development. Organizations may create a more thorough and successful cybersecurity plan that can recognize and address threats in real time by integrating these technologies. The creation of predictive analytics models that can recognize and react to new risks, including those posed by new varieties of malware or cyberattacks, is another field for future study. Organizations can make sure they are constantly one step ahead of hackers and other dangerous actors by staying ahead of the curve in this way.

Predictive analytics has the potential to revolutionize the cybersecurity sector by giving businesses the means to defend their networks and data against online assaults. The advantages of utilizing predictive analytics in cybersecurity operations are obvious, even though there are still difficulties and ethical issues to be resolved. As a result, in the years to come, we may anticipate continuing development and uptake of this technology.

References

[1]Weiss, M. (2022). The rise of cybersecurity warriors. *Small Wars & Insurgencies*, *33*(1-2), 272-293.

[2]Jasper, S. E. (2017). US cyber threat intelligence sharing frameworks. *International Journal of Intelligence and Counterintelligence*, *30*(1), 53-65.

[3]Ganin, A. A., Quach, P., Panwar, M., Collier, Z. A., Keisler, J. M., Marchese, D., & Linkov, I. (2020). Multicriteria decision framework for cybersecurity risk assessment and management. *Risk Analysis*, *40*(1), 183-199.

[4]Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S. & Upadhyay, N. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work, and life. *International journal of information management*, *55*, 102211.

[5]Angafor, G. N., Yevseyeva, I., & He, Y. (2020). Game- based learning: A review of tabletop exercises for cybersecurity incident response training. *Security and privacy*, *3*(6), e126.

[6]Zhao, S., Blaabjerg, F., & Wang, H. (2020). An overview of artificial intelligence applications for power electronics. *IEEE Transactions on Power Electronics*, *36*(4), 4633-4658.

[7]Obaid, L. (2020). *Business-IT Alignment in US-Based Global Organizations: A Qualitative Exploratory Case Study* (Doctoral dissertation, University of Phoenix).

[8]Hoyt, R. E., Snider, D. H., Thompson, C. J., & Mantravadi, S. (2016). IBM Watson Analytics: automating visualization, descriptive, and predictive statistics. *JMIR public health and surveillance*, *2*(2), e5810.

[9]Chornous, G. O., & Gura, V. L. (2020). Integration of information systems for predictive workforce analytics: Models, synergy, security of entrepreneurship. *European Journal of Sustainable Development*, *9*(1), 83-83.

[10]Hoyt, R. E., Snider, D. H., Thompson, C. J., & Mantravadi, S. (2016). IBM Watson Analytics: automating visualization, descriptive, and predictive statistics. *JMIR public health and surveillance*, *2*(2), e5810.

[11]Srivastava, S., Bisht, A., & Narayan, N. (2017). Safety and security in smart cities using artificial intelligence—A review. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence* (pp. 130-133). IEEE.

[12]Orans, L. H., D'Hoinne, J., & Chessman, J. (2020). Market Guide for Network Detection and Response.

[13]Whitworth, V. (2017). *Swimming with Seals*. Bloomsbury Publishing.

[14]Darktrace, D. (2018). Enterprise Immune System.

[15]Cooke, P. (2021). From the machine learning region to the deep learning region: Tesla, DarkTrace and DeepMind as internationalized local to global cluster firms. In *The Globalization of Regional Clusters* (pp. 33-57). Edward Elgar Publishing.

[16]Darktrace, D. (2018). Enterprise Immune System.

[16]Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, *51*(12), 3448-3470.

[17]Hristov, M., Nenova, M., Iliev, G., & Avresky, D. (2021, November). Integration of Splunk Enterprise SIEM for DDoS Attack Detection in IoT. In *2021 IEEE 20th International Symposium on Network Computing and Applications (NCA)* (pp. 1-5). IEEE.

[18]Barak, I. (2020). Critical infrastructure under attack: lessons from a honeypot. *Network Security*, *2020*(9), 16-17.

[19]Vatis, M. A. (2001). *Cyber-attacks during the war on terrorism: A predictive analysis*. Dartmouth College, Hanover NH Institute for Security.

[20]Oo, M. C. M., & Thein, T. (2022). An efficient predictive analytics system for high dimensional big data. *Journal of King Saud University-Computer and Information Sciences*, *34*(1), 1521-1532.

[21]Awodiji, T. O., & Williamsburg, K. Y. (2022). Malicious Malware Detection Using Machine Learning Perspectives.

[22]Yeboah-Ofori, A., Islam, S., Lee, S. W., Shamszaman, Z. U., Muhammad, K., Altaf, M., & Al-Rakhami, M. S. (2021). Cyber threat predictive analytics for improving cyber supply chain security. *IEEE Access*, *9*, 94318-94337.

[23]Eastman, R., Versace, M., & Webber, A. (2015). Big data and predictive analytics: on the cybersecurity front line. *IDC Whitepaper, February*.

[24]Addae, J. H., Sun, X., Towey, D., & Radenkovic, M. (2019). Exploring user behavioral data for adaptive cybersecurity. *User Modeling and User-Adapted Interaction*, *29*, 701-750.

[25]Gulati, H. (2015, March). Predictive analytics using data mining technique. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 713-716). IEEE.

[26]Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of retailing*, *93*(1), 79-95.

[27]Awodiji, T. O. (2021). Interactive Dashboard Design for Manager, Data Analyst and Data Scientist Perspective. 11(19).

[28]Kent, A. D. (2016). Cyber security data sources for dynamic network research. In *Dynamic Networks and Cyber-Security* (pp. 37-65).

[29]Iannacone, M., Bohn, S., Nakamura, G., Gerth, J., Huffer, K., Bridges, R., & Goodall, J. (2015, April). Developing an ontology for cyber security knowledge graphs. In *Proceedings of the 10th Annual Cyber and Information Security Research Conference* (pp. 1-4).

[30]Bey, Z. T., & Agyeman, M. O. (2023, January). An Analysis of Cybersecurity Data Breach in the State of California. In *Cybersecurity in the Age of Smart Societies: Proceedings of the 14th International Conference on Global Security, Safety and Sustainability, London, September 2022* (pp. 159-170). Cham: Springer International Publishing.

[31]Adahman, Z., Malik, A. W., & Anwar, Z. (2022). An analysis of zero-trust architecture and its cost-effectiveness for organizational security. *Computers & Security*, *122*, 102911.

[32]Song, W., Li, X., Afroz, S., Garg, D., Kuznetsov, D., & Yin, H. (2020). Automatic generation of adversarial examples for interpreting malware classifiers. *arXiv preprint arXiv:2003.03100*.

[33]Kandasamy, K., Srinivas, S., Achuthan, K., & Rangan, V. P. (2022). Digital Healthcare-Cyberattacks in Asian Organizations: An Analysis of Vulnerabilities, Risks, NIST Perspectives, and Recommendations. *IEEE Access*, *10*, 12345-12364.

[34]Ponsard, C., Massonet, P., Grandclaudon, J., & Point, N. (2020, September). From lightweight cybersecurity assessment to SME certification scheme in Belgium. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 75-78). IEEE.

[35]Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Ahmad Khan, R. (2020, May). Healthcare data breaches: insights and implications. In *Healthcare* (Vol. 8, No. 2, p. 133). MDPI.

[36]Argaw, S. T., Troncoso-Pastoriza, J. R., Lacey, D., Florin, M. V., Calcavecchia, F., Anderson, D., & Flahault, A. (2020). Cybersecurity of Hospitals: discussing the challenges and working towards mitigating the risks. *BMC medical informatics and decision making*, *20*, 1-10.

[37]White, S., & Dastidar, P. (2021). Lockheed Martin Acquisitions: stay the course or change strategy? *The CASE Journal*, *17*(4), 494-541.

[38]Savola, R. M. (2017, September). Current level of cybersecurity competence and future development: case Finland. In *Proceedings of the 11th European Conference on Software Architecture: Companion Proceedings* (pp. 121-124).

[39]Shackleford, D. (2017). Cyber threat intelligence uses successes and failures: The sans 2017 CTI Survey. *SANS Institute*.

[40]Nadeem, M. (2020). New Normal Chief Marketing Officer (CMO): Branding cybersecurity as a return on trust beyond pandemic. *British Journal of Marketing Studies*, *8*(5), 31-64.

[41]Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, *1*(12), 557-560.

[42]Daimi, K., & Peoples, C. (Eds.). (2021). *Advances in Cybersecurity Management*. Springer.

[43]Pan, Z., Sheldon, J., Sudusinghe, C., Charles, S., & Mishra, P. (2021, February). Hardware-assisted malware detection using machine learning. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 1775-1780). IEEE.

[44]Ustundag, A., Cevikcan, E., Ervural, B. C., & Ervural, B. (2018). Overview of cyber security in the industry 4.0 era. *Industry 4.0: managing the digital transformation*, 267-284.

[45]Lamsal, B. (2020). *Cloud-based Cybersecurity Products, a Detail Analysis, and the Awareness Platform* (Doctoral dissertation, Utica College).

[46]Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, *80*(5), 973-993.

[47]Linda, O., Manic, M., & Vollmer, T. (2012, August). Improving cyber-security of smart grid systems via anomaly detection and linguistic domain knowledge. In *2012 5th International Symposium on Resilient Control Systems* (pp. 48-54). IEEE.

[48]Esteves, J., Ramalho, E., & De Haro, G. (2017). To improve cybersecurity, think like a hacker. *MIT Sloan Management Review*.

[49]Mahmood, T., & Afzal, U. (2013, December). Security analytics: big data analytics for cybersecurity: A review of trends, techniques, and tools. In *2013 2nd national conference on Information assurance (ncia)* (pp. 129-134). IEEE.