

Article

Predictive Analytics and Machine Learning for Cyber Risk Assessment in Critical Information Systems: A Framework for Proactive Threat Identification and Risk Mitigation

Mohd Jahidul Hoque^{1*}, Abdul Rahman Mohammad², Mahbub Hasan³, Muhammad Farhan Nafees⁴

1. BSc in Computer Science & Engineering, Southeast University, Dhaka, Bangladesh
2. Bachelor in Technology, Jawaharlal Nehru Technological University, Hyderabad
3. Master of science in information studies Trine University, USA
4. Bachelors in Business Administration, Shah abdul latif university, Pakistan

* Correspondence: jhaq03@gmail.com¹, mahbub0016@gmail.com³, farannafees9@gmail.com⁴

Abstract: Based on historical data of cyber-attacks on critical information systems, this research examines the possibility of using predictive analytics and machine learning as proactive cyber risk assessment tools. The research uses exploratory data analysis to explore the extent of key types of cyberattacks, patterns of cyber escalation over time, regional differences in cyber risk distribution and the relationship between mitigation technologies and new threats. Results show that a limited number of attack types, including Malware, DDoS and Phishing attacks, account for most of the attacks worldwide, indicating the need to prioritize defensive resources. The temporal analysis also shows that the attacks have consistently increased over the years, culminating in a dramatic increase over the last few years, highlighting the predictive power of historical data in forecasting future attacks. The findings of geospatial analysis revealed inequitable distribution of cyber-risk, with technologically advanced countries having more cyber-attacks that are more frequent, and the emphasis on mitigation technologies also focused on vulnerability management and new solutions like blockchain and encryption. Predictive analytics and machine learning offer a methodological approach to pattern recognition, risk factor identification, and the ability to proactively identify threats. This framework is an improvement on cybersecurity planning because it integrates predictive threat detection and proactive mitigation measures, providing a structured way for organizations to plan for threats and allocate resources appropriately. The results will advance academic research and practical cybersecurity solutions by highlighting the potential of leveraging historical cyberattack data to translate it into actionable intelligence for proactive risk management in critical information systems.

Citation: Hoque, M. J., Mohammad, A. R., Hasan, M & Nafees, M. F. Predictive Analytics and Machine Learning for Cyber Risk Assessment in Critical Information Systems: A Framework for Proactive Threat Identification and Risk Mitigation. American Journal of Economics and Business Management 2023, 6(3), 229-241

Received: 03rd Dec 2022

Revised: 18th Jan 2023

Accepted: 05th Feb 2023

Published: 26th Mar 2023



Copyright: © 2023 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

Keywords: Cyber-attacks, Information, Predictive Analytics, Machine Learning, Cyber Risk, Data

1. Introduction

1.1 Background of the Study

The growing use of digital platforms among people across the globe created an opportunity for industry to collect customer data that helps in taking business-oriented decisions. These platforms facilitate the organizations to collect essential data such as location tracking and personally identifiable information [1]. This information gives insight to a service provider into the requirement of the customer in a specific manner that

further makes it capable of overcoming the pain points or unmet need of the customer. Companies are supposed to behave responsibly toward the collected personal data of an individual by constraining it to the periphery of transformation of business. Launching any new product and services in the market requires effective ideas and the aforementioned collected data of customers from digital platforms become a good source for enterprises. According to the report of McKinsey, around \$300 billion is generated from online promotion of products and services at the global scale, which conveys the strength of customer data in a business environment [1]. The same report also highlighted the other side aspects by outlining the factor of cybersecurity across the digital platforms, where a sensitive approach of people has been observed toward the data privacy breaches on online platforms.

1.2 Problem Statement

The McKinsey report depicted the growing concern of people by describing consumer cautious behavior or approach to share the data on digital platforms of any service provider [1]. This aforementioned change in the consumer behavior shaped a mental status where an individual is not ready to share one's personal data to its service provider, which is actually useful for business strategy. A survey was conducted by the authority of McKinsey among 1000 customers belonging to North America and it was found that most of the consumers were aware of the high-profile data breach that took place in the last few years [1]. Alternatively, the cyberthreat is gradually becoming a well-known incident among common people, which further provoked the customer to become cautious about its personal data sharing on digital platforms.

1.3 Aim of the Research

This research aims to leverage predictive analytics and machine learning capabilities on historical cyberattack trends to evaluate cyber risks, proactively detect threats and facilitate sound cyber mitigation in critical information systems.

1.4 Research Objectives

The research objectives are given below:

- a. To examine the trends in cyberattacks, major cyberattack types and compare their prevalence in cyberattack data.
- b. To understand trends and temporal and geographical patterns in cyberattack activity to detect significant trends in cyber-risk.
- c. Use predictive analytics and machine learning to detect patterns and factors related to cyber-risk.
- d. To study cybersecurity mitigation technologies and to create a blueprint of a proactive approach for threat identification and risk mitigation.

1.5 Research Questions

Question 1. What are the fundamental types of cyberattacks and how common are they in the past history of cyberattacks?

Question 2. What are the temporal and geographical characteristics of Cyber Attack activity and Cyber Risk?

Question 3. What are the key patterns and factors cyber-risk can be detected from predictive analytics and machine learning?

Question 4. What are some ways in which cybersecurity mitigation technologies can help identify threats and mitigate risk proactively in critical information systems?

1.6 Significance of the Study

This study is impactful since the number and variety of cyber-attacks are growing and organizations with critical information systems are facing significant challenges. While traditional Cyber security strategies focus on response to Cyber threats, predictive Analytics and predictive Cyber security techniques provide opportunities to detect patterns of Cyber threats that may foreshadow their potential escalation. This study uses historical data of cyberattacks over the various types of attacks, locations, and technologies employed to mitigate cyber risks, to help understand the data-driven

approach to cyber-risk assessment. The study can help cybersecurity practitioners and organizations detect key attack patterns, gain insights into the temporal and geographical shifts in attacks and determine what factors contribute to high cyber-risk. Moreover, the proposed framework ties predictive threat identification to proactive threat mitigation, which helps with better cybersecurity planning and risk prioritization. The results can also be used for academic research with the goal of showing the use case of historical cyber-attack data for predictive cyber security analysis and to give a framework for using machine learning techniques for proactive cyber risk assessment in critical information systems.

1.7 Evaluation of Cyber Risk in Information Systems

Evaluating cyber risk in information systems is not simply about identifying specific vulnerabilities and weaknesses, it is about the complete assessment of assets, threats, vulnerabilities, impact and countermeasures. Landoll [2] notes that "A security risk assessment delivers structured information to the manager for understanding threats to the organization's assets, identifying vulnerabilities in the existing controls and choosing appropriate safeguards. According to his framework, assessment is a process that consists of project definition and preparation, data collection, risk analysis, risk mitigation and reporting. This method provides a necessary starting point since risk assessment is not of value unless it is linked to organizational decision making. While its assessment approach also suggests that cyber risk evaluation relies heavily on the quality and quantity of information gathered in the assessment process, making it difficult when threat conditions are rapidly evolving.

Moreover, complexity of modern information systems makes it difficult to consider approaches based on the isolated technical indicators of cyber risk. Ganin et al. [3] stated that cybersecurity risk assessment is made up of several criteria and conflicting concerns that need a structure that could accommodate a range of cyber risk dimensions for decision making. This multicriteria view is important as Security decisions may be based on various threats, characteristics of the system, vulnerabilities and consequences that cannot be captured with a single risk measure. The framework thus focuses not just on identifying vulnerabilities, but on supporting prioritization and management decisions. From this point of view, the critical information systems need to take into account interconnected situations instead of considering the security events as separate events.

Finally, the evaluation of cyber risk needs a journey from cyber risk factors to the significance of those factors for an organization's security. With reference to Amin [4] a pragmatic approach to cyber risk assessment, it is important to structure the process of the assessment so that cyber risk can be assessed in a way instead of in the form of disparate security observations. This view reinforces the traditional risk assessment approach, focusing more on the applicability in the operational context and structured evaluation. However, a process-oriented evaluation can still be reliant on the availability of and interpretation of relevant risk data. Therefore, the capacity to discern meaningful patterns within the underlying evidence base is closely related to the effectiveness of cyber risk evaluation, and should not be assumed to be a given.

This is especially if the system components in a distributed information system are interconnected and security conditions differ. Palko et al. [5] argue that cyber security risk modelling in distributed information systems must be able to deal with the complexity, size and distribution of these systems. Their work shows that it is important to consider interactions between the parts of a system, not just limited security observations. This modelling view complements the assessment as it establishes the grounds for modelling cyber risk situations in a structured manner. Distributed environments continue to be complex, which further underscores the need for an analytical approach that can process more information and more varied information in a distributed environment, especially when the goal is to detect evolving patterns of threats and make timely risk decisions based on that information.

One of the other limitations is the lack of reliable cyber risk data that can be used to inform the development of effective assessment and modelling efforts. When Cremer et al. [6] ally reviewed the cybersecurity literature, they reviewed 79 unique datasets out of the wealth of literature reviewed and found a great deal of cyber risk data that is not publicly available. The limitation is relevant for predictive cyber risk assessment as analytical and machine learning approaches need well-structured historical data to be able to detect patterns and give reliable inference. The literature review results indicate that a comprehensive evaluation of cyber risk, along with multicriteria decision making, practical modelling and available evidence, are all essential. The lack of available suitable historical data, however, also compounds the need for data driven approaches to generate temporal, geographical and attack-related patterns from available cyber risk datasets for more proactive threat identification and risk mitigation.

1.8 Assessment of Cyberattack Trends and Threat Intelligence

Cyber threat intelligence has evolved from merely gathering security data to an analytical process to grasp evolving cyber threat trends and enable proactive defense. Cyber threat intelligence is a formalized way of converting the intelligence of cyber threats into knowledge which can be used in decision-making processes for security, especially in the growing connected culture like cloud computing and IoT. The conceptualization is significant because raw threat information is not valuable on its own, without processes to interpret and contextualize it. The digital infrastructures also generate more data, more diverse and faster in speed, about potential threats, however, and traditional information processing methods are beginning to be insufficient for recognizing relevant attack patterns. This makes it essential to have analytical methods that can be used to structure observations from the past and identify meaningful threats as opposed to normal variations.

When cyberattacks are analyzed as trends instead of just events, the analytical aspect of threat intelligence becomes more critical. Samtani et al. [7] mentioned that cyber threat intelligence is a valuable source for cybersecurity industry knowledge since the information of threats can be used for identification of adversarial activities, emerging threats and security requirements. They see threat intelligence as more than just a technical task and link threat intelligence production with other cybersecurity activities and decision making. It means the understanding of cyberattack trends should take into account the source, interpretation and translation of threat information into actionable security knowledge. However, the value of intelligence hinges on the ability to find patterns that are relevant to the tasks at hand, from increasingly diverse sources of information, and so analysis and pattern detection methods are required that are able to process evidence from the past to extract patterns, to build a pattern bank, and to compare past with present. There is a need, however, for analysis and pattern detection methods that can process historical information and extract patterns to create a pattern bank and compare past patterns with current patterns, to enable the usefulness of intelligence.

In recent years, automated intelligence mining has gained more and more attention to tackle this analytical problem. Sun et al. [8] suggest that cyber threat intelligence mining could be of great value for proactive cybersecurity defense by extracting useful knowledge from vast amounts of cyber threat information. They also point to the evolution of analytical techniques that enable the identification of threat entities, relationships, behavior, and patterns, which is the next step in the evolution of cybersecurity intelligence toward more proactive uses. This is especially important for cyberattack trend assessment as historical observations are a way to gather information on repeated attack patterns and evolving attributes of threats. But the quality and diversity of threat information and their availability can affect the effectiveness of intelligence mining, and analytical tools and methods that rely on automation need the right kind of data to produce credible intelligence.

As such, the advancement of cyber threat intelligence relies on its ability to integrate new technologies and adapt to evolving threat landscapes. As reported by Cinar [9] current advancements in cyber threat intelligence are pointing to automation, artificial intelligence, machine learning and other technologies that will enhance cyber threat information collection, analysis and interpretation. This is a forward-looking approach that will further bridge the gap between threat intelligence and predictive cybersecurity as machine learning can help uncover patterns that might not be readily visible using traditional methods of analysis. Meanwhile, the nature of cyber threats is changing, making it difficult to rely on static threat indicators for intelligence systems, which must constantly adapt. Taken together, the literature shows that understanding the trends of cyberattacks effectively depends on the ability of the literature to transform the historical information about threats into contextually relevant and analysis-enabling intelligence. This helps to establish temporal and geographical patterns based on historic cyber-attacks data and helps to lay the groundwork for predictive analytics to make proactive threat identification and cyber risk assessment more robust.

1.9 Application of Predictive Analytics for Cybersecurity

With the volume of security data growing, predictive analytics has emerged as an essential analytical tool in cybersecurity because it is able to identify patterns that can be detected before they result in significant harm to the organization. While traditional cybersecurity analysis focuses on detecting incidents that have already happened, predictive cybersecurity analysis involves using historical and contextual data to detect indicators that are linked to threats. Karamchand [10] equates automation in cybersecurity with the use of machine learning and predictive analytics, which they believe can enhance threat identification and management. This orientation is important as it allows security systems to deal with massive amounts of data efficiently, and automation can help limit the need for manual security monitoring. Yet, predictive automation is still subject to the quality of the data available and the accuracy of the analytical models, requiring suitable preparation of data and the building of the analytical models.

The analytical power of machine learning also enhances predictive cybersecurity by allowing security information to be analyzed using automated pattern recognition and intelligent data analysis. As pointed out by Sarker [11] machine learning helps in cybersecurity by providing intelligent data analysis, automation and creating techniques that can be used to meet newer security needs. This view reveals that machine learning can not only be applied to a specific cybersecurity function, but it can also be used in multiple phases of a security analysis, such as detecting abnormal behavior and also helping automated selection processes. While this is a wider applicability, it also poses an important methodological issue due to the different nature of cybersecurity problems, with varying data characteristics and analysis needs. Thus, predictive cybersecurity can't be solely based on machine learning; it needs to be combined with the proper data and analytical aim, as well as the selected machine learning method.

Predictive analytics is especially useful in cloud-based environments with distributed resources and interconnected services, creating a challenge for cybersecurity monitoring. Bhardwaj and Kaushik [12] proposed predictive analytics based cyber security framework which can assist in cloud security evaluation by utilizing their predictive capabilities to detect potential security risks. This framework is oriented towards perspective and it's applicable here because in cloud environments, there's a lot of operational and security information generated which can be used for predictive analysis. It also suggests that the infrastructure in which threats show up should be taken into account in predictive cybersecurity and not a one-size-fits-all analysis approach. However, the success of predictive models in cloud environments with constantly changing threat and security information relies on their ability to allow for the shifting threat conditions and the variety of information, thus making adaptable analytical processes a central requirement.

Predictive analytics is not limited to the security infrastructure of any single organization, however, but can be used in broader supply chain environments where poor security in one organization can ripple through others. Cyber threat predictive analytics is used to enhance the security of the cyber supply chain as described by Yeboah-Ofori et al. [13] to anticipate and assess threats in supply chain relationships. This not only expands the use of predictive cybersecurity, but also shifts its focus from system protection to risk assessment in inter-connected environments. The study further confirms that historical threat information is relevant in determining patterns that can be used to identify future cyber threats. But as systems become more complex and interdependent, so do predictive methods need increasingly rich amounts of data to capture complex threat conditions and relationships. All these studies underscore predictive analytics and machine learning's capacity to bolster cybersecurity, turning past security data into actionable forecasts and simplifying analytical tasks and enabling earlier detection of threats. It offers a methodological basis for the use of predictive analysis on historical cyberattacks' trends in the present study, where temporal, geographic and attack related information can be analysed to help in the proactive cyber risk assessment and mitigation.

1.10 Machine Learning for Cyber Threat Prediction and Risk Assessment

The importance of machine learning for cyber threat prediction has grown in relevance in the past few years, as traditional risk assessment methods may struggle to effectively address security challenges where there are numerous connected assets, vulnerabilities, and threat conditions, such as the case with critical information systems. In the light of the work of Kure et al. [14] an integrated framework of cybersecurity risk management is a hybrid of risk management and risk prediction, which enhances protection of critical infrastructure. They're making a meaningful contribution because they're putting predictions not just in the detection box, but in the context of risk management. This integration allows cybersecurity assessment to shift from detection of existing vulnerabilities to predicting potential risk situations. The complexity of critical infrastructure, however, suggests that multiple interrelated security factors should be considered for predictive risk assessment, and that this may make model development and interpretation difficult. The study highlights the need for predictive capabilities and a structured risk management process when evaluating cyber threats in critical environments.

The predictive aspect of machine learning is better manifested when a prediction is linked to threat detection and risk assessment as early as possible. As noted by Maddireddy and Maddireddy [15], Artificial intelligence can aid proactive cyber defense by enabling the earlier detection of threats, as well as helping to evaluate the risks involved. This view is crucial because the earlier a threat is detected, the more time security stakeholders will have to respond before it becomes a major incident. AI thus transforms cybersecurity from a more reactive approach into what can be considered predictive security management. However, early prediction is of any value only if analytical systems can provide indications of potential threats that are adequate in quality and can be converted into useful risk information. This motivates the use of machine learning in cyber risk assessment to also take the practical meaning of the output of the model into account, as well as its level of predictive accuracy.

The usefulness of predictive risk assessment is also proven in complex urban settings where digital infrastructures, connected services and different technologies co-exist. Kalinin et al. [16] argue that the key to the cybersecurity risk assessment for smart city infrastructures is to take into account the complexity of components and their interactions. Their work emphasizes the need to view cybersecurity risks at an infrastructure level, rather than to analyze individual technologies separately. This is especially true of machine learning as predictive models can analyze several risk factors, and find patterns where these patterns might be difficult to evaluate based on individual observations. But, due to the interdependence of smart city infrastructure, there are some challenges in terms

of data heterogeneity and varying risk conditions. Overall, the reviewed studies suggest that machine learning and AI can play a role in the cyber threat prediction process by bringing together several risk indicators (RI) and assisting in early detection of threats and enhancing risk assessment processes in information infrastructures. The results of this study form the basis for applying the historical information about cyberattacks to the current study, where the observed patterns of attacks can be analyzed to look for factors that are correlated with cyber risk and thus enable a proactive approach to threat identification and mitigation.

1.11 Literature Gap

The literature examined confirms the significance of the cyber risk assessment, threat intelligence, predictive analytics, and machine learning in cybersecurity's proactive approach. There is, however, little focus on combining historical cyberattacks trends, temporal and geographical trends, and mitigating technologies into a single predictive cyber risk assessment model of critical information systems.

2. Materials and Methods

2.1 Research Design

In this research, the quantitative research design (QRD) was used which integrates exploratory data analysis (EDA) and machine learning in order to analyze historical cyber-attacks patterns and to assist in predictive cyber risk assessment. The study uses a data-driven analytical approach, in which prevalent cyberattack observations from the past are analyzed to capture the main types of attacks, temporal patterns, geographic differences, and links with cybersecurity mitigation technologies. Exploratory analysis is employed to describe the distribution and characteristics of cyberattacks, and machine learning has been integrated to detect patterns and factors related to cyber risk. The research then analyzes the obtained results from the analytical perspective with the purpose to ensure proactive threat identification and risk mitigation in critical information systems. This design offers a linear flow of evidence from the past cyberattacks to predictive analysis and paves the way for a framework that links the cyber risk identification to proactive cybersecurity decision making.

2.2 Data Source and Description

The data for this study is from the Birkbeck_Cyber_Trends dataset from Mendeley Data published on 14 November 2023 [17]. The dataset is available in CSV format in the file CT-0711-1222.csv, and includes historical data on cyberattack trends in 36 countries, including 42 types of cyberattacks and 98 types of cybersecurity mitigation technologies [17]. The data contains the variables for mention of cyberattacks, geographical data, contextual data (conflicts and public holidays), and mentions of various solutions in cybersecurity. The data share these characteristics, which make it well suited for study of prevalence of cyberattacks, temporal and geographic patterns of cyberattack activity, and the connection between cyber threats and cyber mitigation strategies. Thus, the dataset becomes the main empirical material analyzed to explore the phenomenon of predictive cyber risk assessment in this research.

2.3 Data Preprocessing

The Birkbeck_Cyber_Trends data has undergone data preprocessing to ensure it can be reliably explored and predicted. This includes analyzing the data types and structure of the variables, searching for missing or incomplete observations, and verifying for any duplicated records, along with verifying that numerical and categorical variables are presented in an appropriate manner for analysis. Variables that carry temporal information are prepared for trend analysis, and the variables pertaining to attack-related, geographical, contextual and cybersecurity mitigation are grouped by functions in the analysis. If necessary, variables are converted and/or recorded into appropriate formats, but not their meaning. Another aspect of the preprocessing step is the distribution and consistency check of the variables that have been selected to minimize the potential data quality problems and build an appropriate analytical dataset for the exploratory analysis and machine learning.

2.4 Exploratory Data Analysis

Prior to starting a predictive modelling effort, exploratory data analysis is done to discover meaningful patterns and relationships in the historical cyberattack data. This analysis addresses four salient aspects of the data: the relative rates of the major types of cyberattacks, temporal trends in cyberattack activity, geographic variation in cyber risk, and distribution of cybersecurity mitigation technologies relative to emerging threats. Dominant attack categories, trends across time, country differences and trends with the use of mitigation technologies are identified through descriptive analysis and graphical visualization. The four main visualizations created for the Results section are thus intended to give a clear, quick representation of the empirical patterns that are pertinent to the research questions and goals. The results of this step also serve as a foundation for choosing the meaningful variables to be used in the next step (predictive analysis).

2.5 Feature Engineering and Variable Selection

The process of feature engineering and variable selection is performed to find out variables that offer significant information for prediction and evaluation of cyber risk. These variables are analyzed based on their significance in relation to the activity of cyberattacks and their temporal, geographical and cybersecurity mitigation characteristics. Attack data can be summarized into meaningful attack activity indicators and temporal data can be converted to appropriate features for detecting shifts in attack behaviors. Where appropriate based on the characteristics of the data set, variables with limited analytical significance and/or with a high proportion of missing information and/or redundant information may be omitted. The final feature set is identified by exploratory findings and the conceptual relevance to the research goals and suitability for machine learning and ensures that the selected variables contain a clear relationship with cyber risk and proactive threat identification.

2.6 Machine Learning Model Selection and Implementation

Machine learning is used to investigate if there are any historical cyberattack patterns that can be used to support cyber risk identification. The choice of predictive model depends on the structure of the dependent variable, the nature of the available predictors and the analytic goal, which is determined following pre-processing and exploratory data analysis. The selected features are used to train appropriate machine learning techniques, and the available observations are partitioned into a training set and a testing set to evaluate the prediction performance on the latter which has not yet been seen by the machine learning. Modelling involves model training, generation of predictions and comparison of the actual and forecasted results. The implementation aims at revealing high-level patterns in the prediction of attacks, as opposed to just reporting on past attacks, and thus links the empirical data with the goal of proactive cyber threat identification.

2.7 Model Evaluation Metrics

Evaluation of the model is performed to assess how reliable and effective the machine learning approach chosen is. The evaluation metrics are defined based on the final predictive task defined from the dataset and target variable. If the modelling task is a classification, the relevant measures can include accuracy, precision, recall, F1 score, confusion matrix, and ROC area under the ROC curve. If the task of the modelling is numerical prediction, then some of the important measures are mean absolute error, mean squared error, root mean squared error, coefficient of determination. The measures give complementary information about predictive performance, and can be assessed by the model beyond a single accuracy measure. The results from the evaluation are used to interpret the pattern of cyber risks as per the research goal of identifying the meaningful cyber risk patterns for proactive threat assessment.

2.8 Predictive Cyber-Risk Assessment

Predictive cyber risk assessment combines the results of exploratory analysis and from machine learning to detect patterns that can reveal variations in cyber risk. The historical attack prevalence, temporal variation, geographical distribution, and relevant cybersecurity indicators help to interpret the outputs of the predictive model to build a data-driven understanding of cyber risk. Those variables that exhibit significant

predictive contribution are given special consideration as these variables can provide evidence of the conditions that are conducive to increased or increased change in cyberattack activity. Prediction is not a confirmation of future attacks, but rather is a valuable tool that can be used in conjunction with historical relationships and model outputs to help to identify potentially significant risk patterns prior to their occurrence. The method combines predictive analysis with proactive threat identification and lays the groundwork for the analysis of suitable risk mitigation measures in critical information systems.

2.9 Limitations

The study suffers from a number of limitations due to the nature and extent of the data and analysis used. The Birkbeck_Cyber_Trends data set reflects the historical trends of cyberattacks reported and does not necessarily reflect all cyberattacks that have taken place within the countries and periods represented. The observed patterns may be affected by variations in reporting and data availability, as well as the frequency of mentions of specific attacks or cybersecurity technology. The dataset also does not contain information about individual events on the network such as packets – the information is more aggregated, which makes it less possible to look at packets or system level intrusion behavior. Moreover, predictions made with machine learning rely on the quality, relevance, and historical representativeness of the features that are available and are not guaranteed to occur or happen again in the future. These restrictions are taken into account when interpreting results and determining their generalizability.

2.10 Ethical Consideration

The study is based on the Birkbeck_Cyber_Trends dataset, which is publicly available and does not include direct interaction with human participants, collection of personally identifiable information or access to confidential organizational systems. The analysis is carried out for academic research and the data set is used within the context of the research and in keeping with its availability. Data are not altered to show certain results, and analysis methods used are applied throughout the chosen observations and variables. The interpretation of cyberattack patterns is objective and does not explicitly reference any individual or organizations beyond those represented in the dataset, as those responsible for the attacks. Machine learning results are also interpreted with caution as the predictive outputs are not a direct indication of future attacks, they are statistical patterns as learned from past information. Academic referencing is used to ensure that the original dataset and related research is adequately acknowledged.

3. Results

3.1 Analysis of Major Cyberattack Types and Their Relative Prevalence

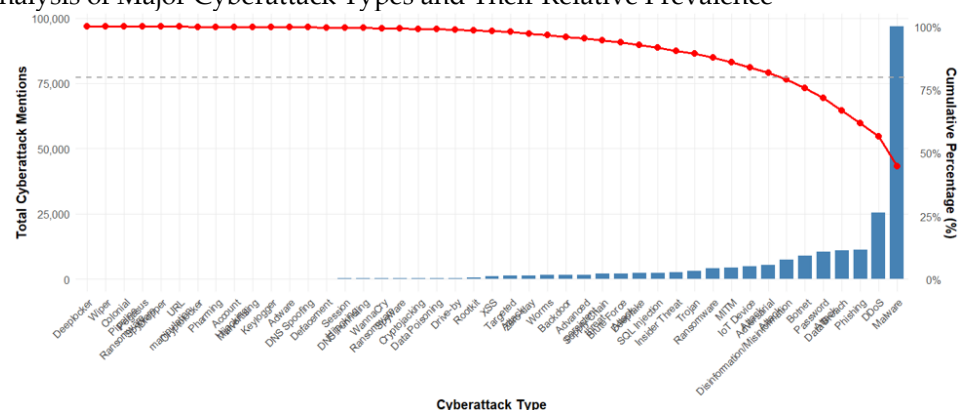


Figure 1. Pareto Analysis of Historical Cyberattack
(Source: Created by learner)

Within the historical data set the Pareto analysis shows that the number of cyber-attack types – including malware, DDoS and phishing – which account for the majority of mentions are limited. This concentration specifies that these attacks are the longest-lasting and most damaging attacks on critical information systems. The visualization helps to

highlight the main categories of attacks affecting the overall cyber-risk pattern and the importance of prioritizing mitigation measures on high-frequency attack categories.

3.2 Temporal Analysis of Cyberattack Trends for Predictive Risk Identification

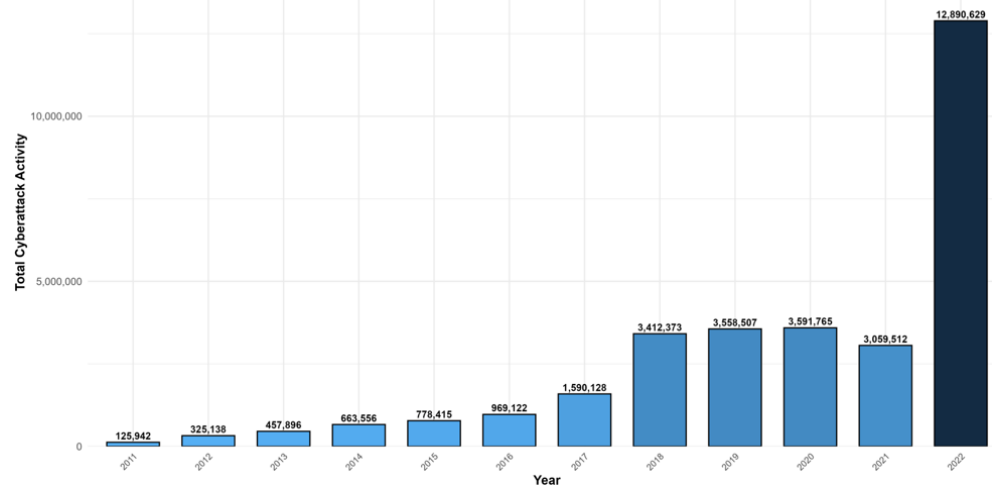


Figure 2. Aggregate Cyberattack Activities Across the Observed Period
(Source: Created by learner)

From 2011 to 2022, there has been a steady increase in the number of cyberattacks, with a significant spike in 2022. This upward trajectory signifies the intensifying global cyber-risk environment and validates the research aim of detecting temporal patterns for predictive risk identification. It reflects the wisdom gained from the past in predicting future escalation of threats and the proactive planning of critical information systems to ensure their resilience.

3.3 Geographical Distribution of Cyberattack Risk Across Countries



Figure 3. Aggregate Ransomware Activity Across Globe
(Source: Created by learner)

This geospatial mapping of ransomware attacks reveals that the U.S. is the most targeted area, trailed by other nations including Brazil, the United Kingdom, Russia, China and Australia. The same geographical variations are likely to be found if the analysis is extended to all cyberattacks in the data set: technologically advanced and digitally interconnected countries are likely to have more frequent attacks. This pattern is in line with the research question concerning geographical characteristics of cyber-risk,

which is how spatial analysis can help to pinpoint vulnerabilities in regions and prioritize cyber security interventions on the international level.

3.4 Cybersecurity Mitigation Technologies Associated with Emerging Cyber Threats

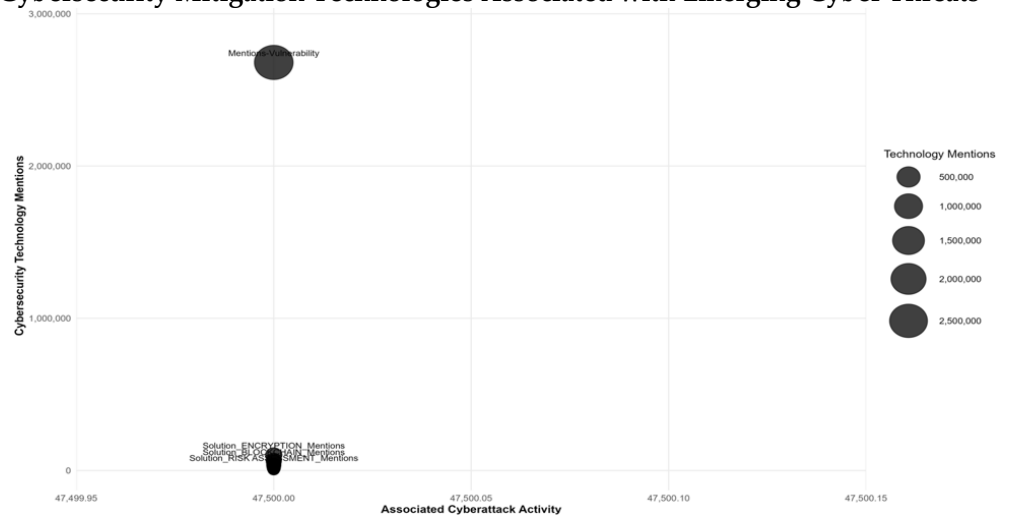


Figure 4. Relationship Between Cyberattack Activity and Mitigation Technology
(Source: Created by learner)

It shows that vulnerability mitigation technologies are the most talked-about technologies in cybersecurity, with the mentions of encryption, blockchain and risk assessment solutions being relatively low compared to others. This distribution is in line with the research goal of connecting with emerging threats and new mitigation technologies, and indicates that vulnerability management continues to play a key role in proactively reducing cyber risk. The visualization highlights the need for building predictive analytics and technology-based solutions for increasing resilience for critical information systems.

4. Discussion

This research indicates that cyberattacks are highly targeted and a small number of categories are responsible for most of the worldwide cyber events, including malware, DDoS, and phishing. This is consistent with the literature, which underscores the resilience of these attacks, as they are so flexible and take advantage of vulnerabilities in the system. The Pareto distribution that was observed in the data set confirms that a very small number of attack types has a disproportionately large impact on cyber-risk and thus defensive resources need to be prioritized accordingly. The results are also a confirmation of the conceptual approach that the most common and most impactful attacks should be taken into account for cyber-risk assessment of critical information systems for efficient risk mitigation.

Temporal analysis suggests a pattern of gradual rising volume of cyberattacks over the years culminating in a significant rise in more recent years, corresponding to the increased level of digital connectivity and attack surfaces. This is in line with the arguments of the literature, which state that complexity and the amount of data driven technologies have increased the cyber-risk exposure. Increasing attack frequency shows the good predictive power of past data for predicting future threat escalation. It also underscores the critical need for predictive analytics and machine learning to detect temporal changes and new trends of future risks that could lead to major cyber events. The findings consequently validate the ability to proactively identify threats by ally analyzing temporal attacks from past attacks.

The geographical and technological analyses also show that the distribution of cyber-risk is not equal between countries and mitigation technologies. Advanced digital infrastructures lead to more attacks, which is in line with the literature and the vulnerability of interconnected digital systems. The visualization of mitigation

technologies reveals that the approach of vulnerability management is still the most discussed and implemented solution, while new solutions like blockchain and encryption are growing in popularity. This is an indication of a changing technological world in which increasingly predictive analytics and machine learning tools are being incorporated into the proactive defense mechanisms. The overall results show that a combination of spatial, temporal and technological dimensions can provide a holistic cyber-risk assessment framework for proactive identification and mitigation of threats.

4.1 Future Work

Future studies should work towards establishing a single predictive model that combines temporal, geographical, and technological aspects of the problem to provide more accurate predictions for proactive cyber risk assessment. Having more complete data, including real-time threat intelligence and geopolitical events and industry-specific vulnerabilities, will enhance the predictive accuracy. The machine learning algorithms can be further optimized to detect intricate relationships between the type of attacks and the type of mitigation. One direction of research that should be pursued in the future is adaptive learning systems that can change over time to evolve with different threats and continually refine predictions of risks. Adding cross-sectoral data from critical infrastructures, like healthcare, finance and energy will bring a more comprehensive view of the systemic cyber-risk and increase the applicability of predictive analytics for global cybersecurity planning.

4.2 Recommendations

Organizations are advised to adopt predictive analytics and machine learning frameworks as means to detect early signs of cyber threats and prioritize the mitigation strategies based on the identified cyber threat. The emphasis for cybersecurity practitioners should shift from using historical data only in conjunction to real-time monitoring to create more comprehensive situational awareness and risk forecasting. Governments and institutions need to have a standardized approach to data sharing and support predictive modelling, enhancing the global threat intelligence. Incorporation of more sophisticated technologies, including blockchain and encryption, into investments to complement vulnerability management and improve the capability of proactive defense of critical information systems.

5. Conclusion

The research work highlights the potential of predictive analytics and machine learning to revolutionize traditional cybersecurity practices and help identify threats and mitigate risks proactively. Examining past cyberattack trends, there are striking temporal trends, geographic variations, and technological connections and relationships that illustrate the dynamic nature of cyber-risk. The incorporation of these dimensions into a comprehensive predictive model can enable organizations to foresee threats and make better use of resources. The research highlights the significance of data-driven decision making for cybersecurity and lays the groundwork for future predictive models that will respond to changing threats. In conclusion, the results provide valuable insights that help develop proactive cybersecurity approaches to protect critical information systems from potential threats in the future.

REFERENCES

- [1] V. Anant, L. Donchak, J. Kaplan, and H. Soller, "The consumer-data opportunity and the privacy imperative," McKinsey & Company, Apr. 27, 2020. [Online]. Available: <https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative>
- [2] D. J. Landoll, *The Security Risk Assessment Handbook: A Complete Guide for Performing Security Risk Assessments*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2021.

- [3] A. A. Ganin, P. Quach, M. Panwar, Z. A. Collier, J. M. Keisler, D. Marchese, and I. Linkov, "Multicriteria decision framework for cybersecurity risk assessment and management," *Risk Analysis*, vol. 40, no. 1, pp. 183–199, 2020, doi: 10.1111/risa.12891.
- [4] Z. H. Amin, "A practical road map for assessing cyber risk," *Journal of Risk Research*, vol. 22, no. 1, pp. 32–43, 2019, doi: 10.1080/13669877.2017.1351467.
- [5] D. Palko, T. Babenko, A. Bigdan, N. Kiktev, T. Hutsol, M. Kuboń, et al., "Cyber security risk modeling in distributed information systems," *Applied Sciences*, vol. 13, no. 4, Art. no. 2393, 2023, doi: 10.3390/app13042393.
- [6] F. Cremer, B. Sheehan, M. Fortmann, A. N. Kia, M. Mullins, F. Murphy, and S. Materne, "Cyber risk and cybersecurity: A review of data availability," *The Geneva Papers on Risk and Insurance Issues and Practice*, vol. 47, no. 3, pp. 698–736, 2022, doi: 10.1057/s41288-022-00266-6.
- [7] S. Samtani, M. Abate, V. Benjamin, and W. Li, "Cybersecurity as an industry: A cyber threat intelligence perspective," in *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, pp. 135–154. Cham, Switzerland: Springer International Publishing, 2020.
- [8] N. Sun, M. Ding, J. Jiang, W. Xu, X. Mo, Y. Tai, and J. Zhang, "Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 3, pp. 1748–1774, 2023.
- [9] B. Cinar, "A study on cyber threat intelligence based on current trends and future perspectives," *Advances and Challenges in Science and Technology*, vol. 5, pp. 37–59, 2023.
- [10] G. K. Karamchand, "Automating cybersecurity with machine learning and predictive analytics," *Journal of Computational Innovation*, vol. 3, no. 1, 2023.
- [11] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects," *Annals of Data Science*, vol. 10, no. 6, pp. 1473–1498, 2023.
- [12] A. Bhardwaj and K. Kaushik, "Predictive analytics-based cybersecurity framework for cloud infrastructure," *International Journal of Cloud Applications and Computing*, vol. 12, no. 1, pp. 1–20, 2022.
- [13] A. Yeboah-Ofori, S. Islam, S. W. Lee, Z. U. Shamszaman, K. Muhammad, M. Altaf, and M. S. Al-Rakhami, "Cyber threat predictive analytics for improving cyber supply chain security," *IEEE Access*, vol. 9, pp. 94318–94337, 2021.
- [14] H. I. Kure, S. Islam, and H. Mouratidis, "An integrated cyber security risk management framework and risk prediction for the critical infrastructure protection," *Neural Computing and Applications*, vol. 34, no. 18, pp. 15241–15271, 2022.
- [15] B. R. Maddireddy and B. R. Maddireddy, "Proactive cyber defense: Utilizing AI for early threat detection and risk assessment," *International Journal of Advanced Engineering Technologies and Innovations*, vol. 1, no. 2, pp. 64–83, 2020.
- [16] M. Kalinin, V. Krundyshev, and P. Zegzhda, "Cybersecurity risk assessment in smart city infrastructures," *Machines*, vol. 9, no. 4, Art. no. 78, 2021.
- [17] Z. J. S. Almahmoud, "Birkbeck_Cyber_Trends," Mendeley Data, Version 1, Nov. 14, 2023. DOI: 10.17632/vt9gj37ykp.1. Available: <https://data.mendeley.com/datasets/vt9gj37ykp/1>