AN INVESTIGATION ON THE FUNCTION OF MACHINE LEARNING IN CYBER DEFENCE AND THE IDENTIFICATION OF ITS CAPABILITIES BEYOND THREAT DETECTION.

Li Qinying[1], Divya Midhunchakkaravarthy[1]

[1]Lincoln University College, Petaling Jaya, Malaysia.

ABSTRACT

Particularly in relation to its role beyond the traditional threat detection, this study investigates the game-changing implications of ML for cybersecurity. In the face of ever-evolving cyber threats, conventional methods often fall short. By studying massive information and spotting patterns, machine learning might substantially enhance cybersecurity methods. This paper focusses on ML techniques that might be useful in several domains, including threat prediction, anomaly detection, and automated response. Investigated are the ways in which ML models can analyse attack trends over time and identify subtle indicators of impending assaults that humans might overlook. The study also explores the use of ML in real-time response systems, which may adjust to new dangers by learning from new data as it happens. In addition to detection and reaction, the paper highlights that ML can automate routine security tasks, enhance threat intelligence, and optimise resource allocation. Incorporating ML into cybersecurity frameworks may help organisations attain more proactive and flexible security postures. One of these approaches is ML-based behavioural analysis, which provides insight into user conduct and calls attention to anomalies that can suggest security flaws. Machine learning's potential to revolutionise cybersecurity processes is explored extensively in the study's last section. Beyond only detecting threats, it demonstrates how technology may provide novel solutions for prediction, response, and overall security management. The findings may pave the way for future ML research and applications, which might lead to more adaptable and secure cybersecurity systems.

Keywords: Machine Learning, Information Security, The Abilities, Vulnerability Recognition.

INTRODUCTION

Even in this dynamic digital landscape, cybersecurity is a top concern for organisations and individuals. Cyber attacks are becoming smarter every day therefore there's a greater need than ever before for innovative protection solutions. Among these technologies, machine learning (ML) stands out as a game-changer, offering significant advancements over older, more traditional methods of threat detection (Amich & Birhanu, 2021).

Along with its more well-known use in threat detection, this study will also examine machine learning's other applications in cybersecurity. The goal of machine learning,

a branch of artificial intelligence, is to program computers to automatically improve themselves via trial and error. The cybersecurity threat detection game has been turned on its head by machine learning. Computers may use it to spot patterns and outliers that might indicate an attack is about to happen. Its value, however, extends far beyond this critical application.

This investigation also intends to look at other important aspects of cybersecurity, such as automated incident response, vulnerability management, and behavioural analytics. The application of ML techniques in cybersecurity allows for improved attack prediction and mitigation, more precise threat information, and better management of security operations. Finally, the study will investigate if ML can aid in the adaption of security measures to emerging dangers and attack routes. Through an exhaustive examination, this paper illuminated the current capabilities of machine learning in cybersecurity and provided insights into the industry-transforming impact of these technologies. The study aims to shed light on the capabilities of ML and contribute to the development of more robust and flexible cybersecurity strategies by analysing the growing roles it plays (Corsini et al., 2021).

BACKGROUND OF THE STUDY

The merging of machine learning with cybersecurity has brought about a tremendous shift in the field of information security. There has been prior focus on the importance of signature-based threat detection techniques in cybersecurity. According to (Alhogail & Alsabih, 2021), this method would have computers detect malicious activities by comparing them to known threat signatures. This approach has failed to keep up with the dynamic nature of cyber threats, despite its efficacy in the past. When machine learning—a kind of artificial intelligence—was introduced, cybersecurity strategies were drastically altered. Machine learning enables computers to learn from their mistakes and uncover previously unseen patterns in data.

The late 20th century saw a proliferation of applications for early machine-learning models that went beyond their original simplicity, thanks to improvements in processing power and data availability. Machine learning's impact on cybersecurity grew in the 2010s, along with the rise in computer power and algorithmic complexity. The use of neural networks and other pattern recognition and predictive analytics tools enabled the identification of both previously identified and newly emerging threats (Emilie et al., 2021). Beyond threat detection, cybersecurity researchers and practitioners started to look at other applications of machine learning. In order to make them more responsive and dynamic, security solutions began to use ML. This included behavioural analytics, vulnerability management tools, and automated incident response systems.

There was clear potential in these advancements for using machine learning to handle threats, as well as to detect and prevent them. In the ever-changing landscape of cybersecurity, the value of machine learning in developing proactive and adaptable security solutions is becoming more recognised. The goal of this study is to contribute to that literature by investigating machine learning's broader applications in cybersecurity, going beyond threat detection and the enhanced capabilities it offers in this domain (Eslam & Zelinka, 2020).

PURPOSE OF THE RESEARCH

Exploring machine learning's impact on cybersecurity beyond its usual use in threat detection is the study's main objective. Examining how machine learning may enhance automated incident response, exposure management, and behavioural analytics, as well as other broader applications of machine learning in cybersecurity, is the aim of this research. The ultimate goal of the study is to provide insight into the potential of machine learning to develop more proactive, flexible, and all-encompassing cybersecurity solutions. Methods for cybersecurity will be strengthened and fortified by this.

LITERATURE REVIEW

The integration of machine learning (ML) into cybersecurity has become a hot issue due to its revolutionary potential. Conventional cybersecurity strategies, which rely heavily on signature-based detection approaches, have been rendered ineffective by the dynamic nature of cyber threats. In this respect, machine learning's ability to filter through massive amounts of data in quest of patterns has made it an indispensable tool (Boenisch et al., 2021).

Multiple research projects have shown that ML has the potential to enhance danger detection abilities. Methods such as anomaly detection and supervised learning were used to identify both existing and emerging threats by analysing patterns in user behaviour and network traffic. These methods successfully identified sophisticated malware and zero-day threats in their testing. Other important aspects of cybersecurity, beyond threat detection, have recently been the focus of new study. Automated incident response systems powered by ML can quickly analyse security incidents and execute pre-defined procedures to reduce risks these solutions use ML algorithms for better event prioritisation and reaction instead of human methods. Vulnerability managers can foresee potential vulnerabilities by applying ML techniques to attack patterns and historical data. These predictive skills have made it possible for organisations to address security vulnerabilities prior to their exploitation. When it comes to cybersecurity, ML has shown promise in some areas, such as behavioural analytics. Machine learning algorithms may analyse patterns of user activity to identify potential insider threats or compromised accounts. This

approach enhances the ability to detect and respond immediately to internal threats (Luis et al., 2020).

To sum up, machine learning has not only made threat detection much better, but it can also automate responses, manage vulnerabilities, and study behaviour. Research shows that using ML's capabilities in cybersecurity requires a thorough approach (Mehdi et al., 2019).

RESEARCH QUESTION

What is the impact of data quality in threat detection?

RESEARCH METHODOLOGY

RESEARCH DESIGN

The quantitative data analysis was conducted using SPSS version 25. The odds ratio and 95% confidence interval were used to ascertain the strength and direction of the statistical link. The researchers developed a statistically significant criterion at $p < 0.05$. A descriptive analysis was performed to determine the key characteristics of the data. Quantitative approaches are often used to evaluate data obtained from surveys, polls, and questionnaires, as well as data modified by computational tools for statistical analysis.

SAMPLING

A convenient sampling technique was applied for the study. The research relied on questionnaires to gather its data. The Rao-soft program determined a sample size of 1463. A total of 1600 questionnaires were distributed; 1557 were returned, and 57 were excluded due to incompleteness. In the end, 1500 questionnaires were used for the research.

DATA AND MEASUREMENT

A questionnaire survey served as the principal tool for data gathering in the study. The survey had two sections: (A) General demographic information and (B) Responses on online and offline channel variables assessed using a 5-point Likert scale. Secondary data was obtained from many sources, mostly on internet databases.

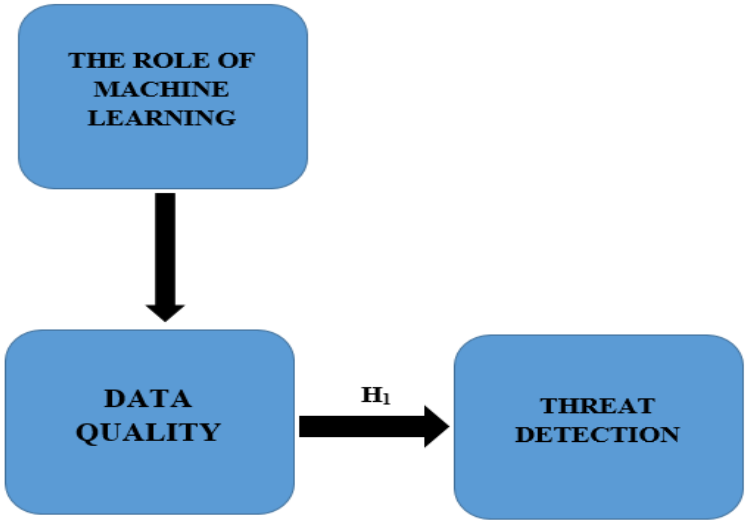
STATISTICAL SOFTWARE

The statistical analysis was conducted using SPSS 25 and MS-Excel.

STATISTICAL TOOLS

To grasp the fundamental character of the data, descriptive analysis was used. The researcher is required to analyze the data using ANOVA.

CONCEPTUAL FRAMEWORK



RESULT

Factor Analysis: One typical use of Factor Analysis (FA) is to verify the existence of latent components in observable data. When there are not easily observable visual or diagnostic markers, it is common practice to utilise regression coefficients to produce ratings. In FA, models are essential for success. Finding mistakes, intrusions, and obvious connections are the aims of modelling. One way to assess datasets produced by multiple regression studies is with the use of the Kaiser-Meyer-Olkin (KMO) Test. They] verify that the model and sample variables are representative. According to the numbers, there is data duplication. When the proportions are less, the data is easier to understand. For KMO, the output is a number between zero and one. If the KMO value is between 0.8 and 1, then the sample size should be enough. These are the permissible boundaries, according to Kaiser: The following are the acceptance criteria set by Kaiser:

A pitiful 0.050 to 0.059, below average 0.60 to 0.69

Middle grades often fall within the range of 0.70-0.79.

With a quality point score ranging from 0.80 to 0.89.

They marvel at the range of 0.90 to 1.00.

Table1: KMO and Bartlett's Test

Testing for KMO and Bartlett's

Sampling Adequacy Measured by Kaiser-Meyer-Olkin .960

The results of Bartlett's test of sphericity are as follows: approx. chi-square

df=190

sig.=.000

This establishes the validity of assertions made only for the purpose of sampling. To ensure the relevance of the correlation matrices, researchers used Bartlett's Test of Sphericity. Kaiser-Meyer-Olkin states that a result of 0.960 indicates that the sample is adequate. The p-value is 0.00, as per Bartlett's sphericity test. A favourable result from Bartlett's sphericity test indicates that the correlation matrix is not an identity matrix.

Table 1: KMO and Bartlett's.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.960
Bartlett's Test of Sphericity	Approx. Chi-Square	3252.968
	df	190
	Sig.	.000

This substantiates that assertions on the execution of a sample are valid. Researchers used Bartlett's Test of Sphericity to evaluate the importance of the correlation matrices. The Kaiser-Meyer-Olkin metric deems the sample adequate when the result is 0.960. According to Bartlett's sphericity test, the p-value is 0.00. The statistically significant findings of Bartlett's sphericity test indicate that the correlation matrix differs from an identity matrix.

INDEPENDENT VARIABLE

The Role of Machine Learning: By empowering systems to autonomously learn from data and improve their performance over time without explicit programming, machine learning (ML) is reshaping several sectors and playing a pivotal role in today's technological environment. Building statistical models and algorithms that enable computers to learn from past data, spot trends, and make predictions is the essence of machine learning. It has a wide range of potential uses in many different fields, including medicine, economics, advertising, NLP, driverless cars, and more. Machine learning algorithms find many applications in healthcare, including image analysis, disease outbreak prediction, and patient-specific therapy planning. One area where ML has proven useful is in the financial sector, namely in the areas of fraud detection, investment portfolio optimization, and trend forecasting. Machine learning's function is twofold: first, to improve the precision and efficacy of current

systems; second, to propel innovation by making possible totally new capacities, like autonomous vehicles or smart virtual assistants. The impact of ML is already substantial, and it will only grow as the technology develops further, revolutionizing how businesses operate, enhancing the quality of decisions made, and creating opportunities for innovation in every area of society. The influence of machine learning is expanding rapidly in both general and niche professional fields, thanks to developments in algorithm design, improvements in computing power, and the availability of massive datasets. In the end, machine learning is going to be a game-changer for tech since it allows for smart systems to evolve, learn, and adapt, which is crucial for moving forward in the digital era (Muna Al & Elena, 2019.

FACTOR

Data Quality: Data quality is defined as the extent to which data satisfies the criteria and requirements for its intended use, and it is an indicator of the data's general condition or integrity. If the researchers want to do good analysis, make good decisions, and run the researcher’s organisation efficiently, the researchers need high-quality data. It covers a lot of ground and has to be right on time, comprehensive, consistent, relevant, and genuine. Data that is accurate is devoid of mistakes and faithfully depicts the things it describes in the actual world. In order to avoid losing crucial pieces of information that might result in inaccurate or incomplete conclusions, completeness is key. The absence of inconsistencies or contradictions that might cast doubt on the veracity of data is what the researchers mean when the researchers talk about consistency across various platforms and sources. To make sure judgements are based on the most recent facts, it's important that the data be up-to-date and accessible when required. In order for data to provide useful insights, it must be relevant, meaning it must match the requirements of the particular job or business activity. The format, kind, or range of data must be consistent with what is anticipated in order for it to be valid for the systems or models that rely on it. Mistaken conclusions, wasted money, faulty analysis, and distrust in data-driven procedures are all possible outcomes of low data quality. Efforts in data governance, data cleansing, validation procedures, and frequent audits are essential to keep data quality good, which is becoming more important as organisations depend on data for strategic decision-making. At its core, data quality goes beyond mere accuracy; it encompasses reliable, accessible, and insight-supporting data that can drive growth, efficiency, and innovation for organisations (Chaudhary et al., 2020).

DEPENDENT VARIABLE

Threat Detection: The term "threat detection" describes the steps used to spot possible dangers to a system's or network's security or harmful actions. It entails keeping an eye on data for signs of intrusion, cyberattacks, or other security breaches using cutting-edge tech, tools, and procedures. An essential part of

cybersecurity is threat detection, which aids organizations' in seeing possible dangers before they do major damage, such as data theft, system damage, or service interruption. Unusual login attempts, malware, or illegal data access are examples of potential dangers that may be identified via the process of analysing network traffic, system logs, user behaviour, and other pertinent data sources. Because these technologies may identify new or emerging dangers that older approaches would overlook, modern threat detection systems often use them to improve detection capabilities. There are two main approaches to threat detection: proactive, which aims to find security holes and reduce risks before an attack happens, and reactive, which deals with security issues as they happen, helping to contain them and get back on the researchers r feet after a breach. Data loss, financial damages, reputational harm, and other negative outcomes may be prevented if organizations can identify threats quickly and accurately. To keep businesses running smoothly and prevent unauthorized access to critical data, it is essential to keep IT infrastructure secure and intact. Keeping ahead of new threats and adjusting detection methodologies to fit the ever-changing cybersecurity environment are essential components of effective threat detection, which is an ongoing process (Giovanni et al., 2021)

Relationship Between Data Quality and Threat Detection: The effectiveness of threat detection systems is highly dependent on the quality of the data they examine, so there is a basic and interdependent link between data quality and threat detection. Accurately detecting and reacting to security threats requires high-quality data (Giovanni et al., 2020). When security systems have access to precise, full, consistent, and up-to-date data, they are better equipped to spot abnormalities or malicious behaviours with more accuracy and less false positives or negatives. To illustrate the point, threat detection systems might miss important security events or fail to identify possible breaches if data on user behaviours, network traffic, or security records is inconsistent, missing, or both. This can cause delays in replies or perhaps the missed breach altogether. Additionally, the system may miss new or changing threats that evade pattern recognition-based classical detection approaches if the data is of low quality. Threat detection algorithms, on the other hand, rely on high-quality data to identify real dangers from fake ones by providing them with the most up-to-date, accurate, and thorough information possible. Organizations need to put money into both advanced threat detection tools and the procedures and plans to keep their data clean because of this correlation. If threat detection systems want to use accurate information, data governance, frequent data cleansing, and validation are musts. Even the most sophisticated detection systems may be rendered useless by low-quality data; these systems can either miss new threats altogether or cause security staff to get overwhelmed by false alerts, which slows down reaction times. In the end, organizations' can better safeguard their assets, infrastructure, and sensitive information against harmful assaults when data quality is good. This is because threat detection becomes more reliable and

efficient. Therefore, in order to construct robust cybersecurity systems that can adjust to new threats and adapt to new problems, there must be a constant emphasis on enhancing data quality (Giuseppina et al., 2021).

On the basis of the above discussion, the researcher formulated the following hypothesis, which was analyse the relationship between Data Quality and Threat Detection.

"H_{01}: There is no significant relationship between Data Quality and Threat Detection."

"H_1: There is a significant relationship between the role Data Quality and Threat Detection."

Table 2: H_1 ANOVA Test.

ANOVA					
Sum					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	65692.704	633	5474.392	3028.874	.000
Within Groups	168.296	866	1.934		
Total	65861.000	1499			

The findings of this investigation are noteworthy. The F value is 3028.874, attaining significance with a p-value of .000, which is below the .05 alpha threshold. This means the **"H_1: There is a significant relationship between the role Data Quality and Threat Detection."** is accepted and the null hypothesis is rejected.

DISCUSSION

Beyond its more common usage in threat detection, machine learning's (ML) revolutionary potential in cybersecurity extends far and wide. By analysing and mitigating security incidents more quickly, ML improves automated incident response by decreasing reaction times and human error. Vulnerability management makes use of machine learning (ML) to examine historical data for potential vulnerabilities. This makes it possible to take preventative measures. Artificial intelligence (AI)-enabled behavioural analytics improve insider threat detection by spotting unusual user actions. Issues with data quality, algorithmic bias, and interoperability persist despite these advancements. It is critical to resolve these issues in order to make the most of ML's impact on all-encompassing cybersecurity policy. Research on machine learning's (ML) impact on cybersecurity has shown its revolutionary potential in several areas, including threat detection and a host of others that strengthen defences against cyberattacks. Machine learning provides scalable and adaptable solutions that can detect and react to changing threats in real-time, in contrast to conventional cybersecurity approaches that often depend

on predetermined rules and signature-based detection. Machine learning (ML) is finding more and more applications beyond threat detection. For example, behavioural analytics is using ML to set baseline activity and identify abnormalities that may suggest account hacks or insider threats. Isolating impacted devices or restricting harmful IP addresses are two examples of how machine learning improves incident response via automation. Systems can respond quickly to identify risks. Organisations may proactively strengthen their defences before assaults happen with the use of predictive analytics driven by ML, which can foresee prospective cyberattacks based on past data. Machine learning algorithms are able to detect and categorise dangerous behaviours rapidly in many domains, including fraud prevention, phishing detection, malware detection, and network traffic analysis. These algorithms often discover novel attack vectors that conventional systems fail to notice. Still, there are obstacles to overcome, most notably in the areas of high-quality data requirements, model interpretability, and protection against adversarial assaults. But this report shows that machine learning is becoming more important for cybersecurity in general, and for threat detection in particular, so it's a vital tool for companies trying to protect their digital surroundings.

CONCLUSION

This research emphasizes the significant contribution of machine learning (ML) to improving cybersecurity, which goes beyond conventional techniques of threat identification. Modern cybersecurity would be incomplete without machine learning, which can sift through mountains of data, spot trends, and adjust to ever-changing dangers. Machine learning has widely acknowledged benefits in threat detection (e.g., anomaly detection, real-time warnings), but it also shines in other important domains like automated incident response, predictive analytics, malware detection, fraud prevention, behavioral analytics, and so on. Together, these features strengthen defenses against both known and unknown threats, and they allow for proactive and adaptive defense against new ones. The use of machine learning has several benefits in enhancing security efficiency, decreasing reaction times, and minimizing risks; nevertheless, there are also obstacles to overcome, such as worries about data quality, the possibility of adversary manipulation, and the need for model openness. In the ever-changing landscape of cyber threats, machine learning is poised to lead the way in cybersecurity innovation. It will help organizations' stay one step ahead of attackers and create digital environments that are stronger and more robust. In order to effectively combat the ever-evolving and complex cyber threat environment, it is crucial to keep incorporating machine learning into cybersecurity measures. By improving automated incident response, fortifying vulnerability management, and refining behavioral analytics, machine learning elevates cybersecurity above traditional threat detection. With these features, more proactive and flexible security measures may be implemented to tackle a broader range of cyber threats. Problems with data quality or system integration must be addressed in order to fully use the potential of ML. To make the most of machine

learning's potential in developing stronger and more efficient cybersecurity solutions, research and development must continue.

REFERENCES

1. Abderrahmen Amich and Birhanu Eshete. 2021. Explanation-guided diagnosis of machine learning evasion attacks. Proceedings of the ACM International Conference on Availability, Reliability and Security Conference.
2. Andrea Corsini, Shanties Yang, and Giovanni Apruzzese. 2021. on the evaluation of sequential machine learning for network intrusion detection. In Proceedings of the International Conference Availability, Reliability, Security.
3. Areej Alhogail and Afrah Alsabih. 2021. Applying machine learning and natural language processing to detect phishing email. Comput. Secur. 110 (2021), 102414.
4. Emilie Bout, Valeria Loscri, and Antoine Gallais. 2021. How machine learning changes the nature of cyberattacks on IoT networks: A survey. IEEE Commun. Surv. Tutor. (2021).
5. Eslam Amer and Ivan Zelinka. 2020. A dynamic Windows malware detection and prediction method based on contextual understanding of API call sequence. Comput. Secur. 92 (2020), 101760.
6. Franziska Boenisch, Verena Battis, Nicolas Buchmann, and Maija Poikela. 2021. “I never thought about securing my machine learning systems”: A study of security and privacy awareness of machine learning practitioners. In Mensch und Computer 2021. 520–546.
7. Giovanni Apprizes, Mauro Andreolini, Luca Ferretti, Mirco Marchetti, and Michele Colajanni. 2021. Modelling realistic adversarial attacks against network intrusion detection systems. ACM Digit. Threats: Res. Pract. (2021).
8. Giovanni Apprizes, Mauro Andreolini, Michele Colajanni, and Mirco Marchetti. 2020. Hardening random forest cyber detectors against adversarial attacks. IEEE Trans. Emerg. Top. Comput. Intell. 4, 4 (2020), 427–439.
9. Giuseppina Andresini, Fergus Pendle bury, Fabio Pierazzi, Corrado Logics, Annalisa Appice, and Lorenzo Cavallaro. 2021. INSOMNIA: Towards concept-drift robustness in network intrusion detection. In Proceedings of the ACM CCS Workshop on Artificial Intelligence and Security.
10. Jacopo Bellasio and Erik Silfversten. 2020. The impact of new and emerging technologies on the cyber threat landscape and their implications for NATO. In Cyber Threats and NATO 2030: Horizon Scanning and Analysis, 88.
11. Luis Dias, Simão Valente, and Miguel Correia. 2020. Go with the flow: Clustering dynamically-defined net flow features for network intrusion detection with DynIDS. In Proceedings of the IEEE 19th International Symposium on Network Computing and Applications (NCA’20). IEEE, 1–10.
12. Mehdi Abagail, Mohammad Pourmahmood Aghababa, and Vahid Solouk. 2019. Heuristic nonlinear regression strategy for detecting phishing websites. Soft Comput. 23, 12 (2019), 4315–4327.

13. Muna Al-Hawawreh and Elena Sitnikova. 2019. Leveraging deep learning models for ransomware detection in the industrial Internet of Things environment. In Proceedings of the IEEE Military Communications and Information Systems Conference.1–6.
14. Sujita Chaudhary, Austin O'Brien, and Shengjie Xu. 2020. Automated post-breach penetration testing through reinforcement learning. In Proceedings of the IEEE Conference on Communications and Network Security (CNS'20). 1–2.