computers

MDPI

*Article*

# Machine Learning Cybersecurity Adoption in Small and Medium Enterprises in Developed Countries

**Nisha Rawindaran [1,2,3], Ambikesh Jayal [4,*] and Edmond Prakash [1]**

[1] Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff CF5 2XJ, Wales, UK; nrawindaran@cardiffmet.ac.uk (N.R.); eprakash@cardiffmet.ac.uk (E.P.)
[2] Aytel Systems Ltd., Cardiff CF3 2PU, Wales, UK
[3] KESS2, Knowledge Economy Skills Scholarships, Supported by European Social Funds (ESF), Bangor University, Bangor, Gwynedd LL57 2DG, Wales, UK
[4] School of Information Systems and Technology, University of Canberra, Bruce, ACT 2617, Australia
* Correspondence: ambi.jayal@canberra.edu.au

**Abstract:** In many developed countries, the usage of artificial intelligence (AI) and machine learning (ML) has become important in paving the future path in how data is managed and secured in the small and medium enterprises (SMEs) sector. SMEs in these developed countries have created their own cyber regimes around AI and ML. This knowledge is tested daily in how these countries' SMEs run their businesses and identify threats and attacks, based on the support structure of the individual country. Based on recent changes to the UK General Data Protection Regulation (GDPR), Brexit, and ISO standards requirements, machine learning cybersecurity (MLCS) adoption in the UK SME market has become prevalent and a good example to lean on, amongst other developed nations. Whilst MLCS has been successfully applied in many applications, including network intrusion detection systems (NIDs) worldwide, there is still a gap in the rate of adoption of MLCS techniques for UK SMEs. Other developed countries such as Spain and Australia also fall into this category, and similarities and differences to MLCS adoptions are discussed. Applications of how MLCS is applied within these SME industries are also explored. The paper investigates, using quantitative and qualitative methods, the challenges to adopting MLCS in the SME ecosystem, and how operations are managed to promote business growth. Much like security guards and policing in the real world, the virtual world is now calling on MLCS techniques to be embedded like secret service covert operations to protect data being distributed by the millions into cyberspace. This paper will use existing global research from multiple disciplines to identify gaps and opportunities for UK SME small business cyber security. This paper will also highlight barriers and reasons for low adoption rates of MLCS in SMEs and compare success stories of larger companies implementing MLCS. The methodology uses structured quantitative and qualitative survey questionnaires, distributed across an extensive participation pool directed to the SMEs' management and technical and non-technical professionals using stratify methods. Based on the analysis and findings, this study reveals that from the primary data obtained, SMEs have the appropriate cybersecurity packages in place but are not fully aware of their potential. Secondary data collection was run in parallel to better understand how these barriers and challenges emerged, and why the rate of adoption of MLCS was very low. The paper draws the conclusion that help through government policies and processes coupled together with collaboration could minimize cyber threats in combatting hackers and malicious actors in trying to stay ahead of the game. These aspirations can be reached by ensuring that those involved have been well trained and understand the importance of communication when applying appropriate safety processes and procedures. This paper also highlights important funding gaps that could help raise cyber security awareness in the form of grants, subsidies, and financial assistance through various public sector policies and training. Lastly, SMEs' lack of understanding of risks and impacts of cybercrime could lead to conflicting messages between cross-company IT and cybersecurity rules. Trying to find the right balance between this risk and impact, versus productivity impact and costs, could lead to UK SMES getting over these hurdles in this cyberspace in the quest for promoting the usage of MLCS. UK and Wales governments can use the research conducted in this paper to inform



**Citation:** Rawindaran, N.; Jayal, A.; Prakash, E. Machine Learning Cybersecurity Adoption in Small and Medium Enterprises in Developed Countries. *Computers* **2021**, *10*, 150. https://doi.org/10.3390/computers10110150

Academic Editor: Paolo Bellavista

Received: 11 September 2021
Accepted: 28 October 2021
Published: 10 November 2021

and adapt their policies to help UK SMEs become more secure from cyber-attacks and compare them to other developed countries also on the same future path.

**Keywords:** UK SME; machine learning cybersecurity; cyber security; machine learning; artificial intelligence; barriers; challenges; cyberspace; cyber awareness

## 1. Introduction

SMEs face a fight for balance when it comes to keeping their data safe and secure. With cyber-attacks rising due to the increase of smart technologies, standard measures are being put in place in line with recent changes to the law, Brexit, UK GDPR, and Cyber Essentials [1] amongst many others. SMEs struggle to understand the bigger concepts of how AI and ML could help. Getting these standards in place requires an intervention to current safety measures of cyber security, and control of varied connections and interactions on the internet.

One solution emerging is the use of MLCS techniques, allowing organizations to identify the cause–effect relationships between breaches and their impact on SMEs. By using statistical techniques reveals characteristic behaviors and patterns for zero-day attacks in cyberspace. Relationships between these anomaly variables can inevitably contribute to the safety of data management within the SME environment. In addition to organizations and varying technologies contributing to the myriad of variables having an effect on cyber security, the human factors also offer a large influence in the security of Internet of Things (IoTs) and devices. Many technical advisory groups offer online contributions to understanding IoTs and how SMEs can cope and live alongside them. Anti-forensic methods, jurisdiction, and service level agreements (SLA) all further aggravate technical, privacy, security, and legal challenges. The presence of GDPR [1] presented by the IoT, and human factors involved, allows for IoTs to be safe and secure within an SME ecosystem.

Industry 4.0 [2] has also contributed to the change in how technology is used. It is with no surprise that the pre-pandemic era was confusing enough for businesses and SMEs to make certain choices and make decisions based on what worked best for their business. Capitalizing on AI and ML would be required to improve capabilities and workflows within the business, inclusive of how their business handled their financial management. In an article written in the International Institute for Management Development (IMD) [3] based in the developed country of Lausanne, Switzerland, IMD focused on Industry 4.0 and COVID-19 and gave thought to the notion that AI was most effective when there was a historical database to learn from and exploit to better predict the future. This is indeed valid; as more historic data or Big Data [4] is collected, there is a better predictability chart to show the "best fit" line. The same article goes on to discuss how perhaps in the past, businesses looked at business drivers and other complex events to drive their business forward, when it could be seen at a simpler angle to succeed. The IMD article discusses how the pandemic created a simpler model to work from in order to survive. Supply and demand took a shift, and its platform and demographics moved to a more resilient virtual environment free from COVID-19, that environment being cyberspace. It is no wonder that the assumption of people's awareness is at stake; indeed, humans now must get to the next level of design with ML and start interacting with machines at a higher rate of speed. Figure 1 below provides a clear representation of the workflow undertaken to give the paper clarity and movement throughout the understanding and knowledge acquired in the presentation of the works.

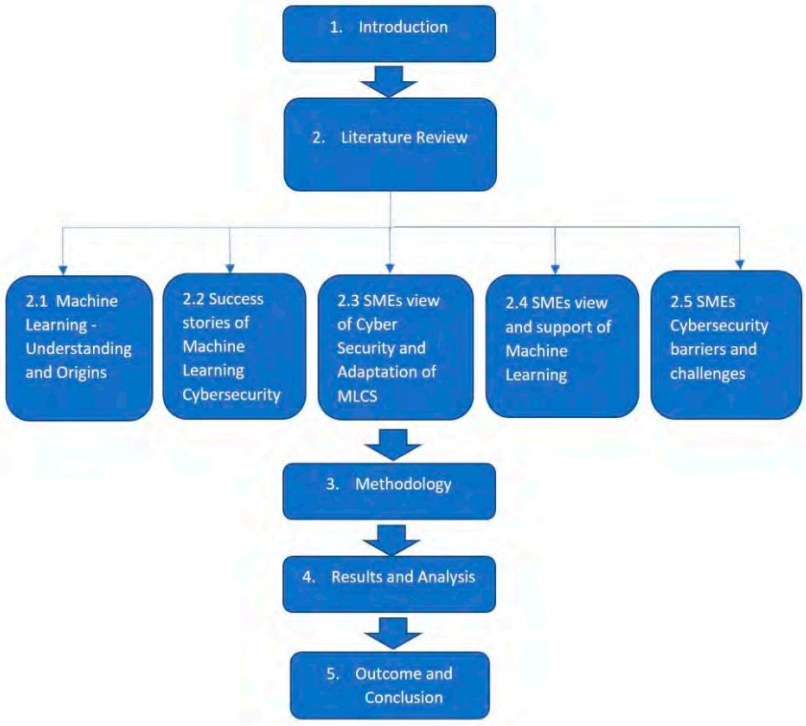


**Figure 1.** Graphical representation of the literature review.

As seen in Figure 1 above, this paper investigates advanced research on the issues of cybersecurity within the SME market in developed nations using machine learning and combining the strength of machine learning and cybersecurity (MLCS). This paper goes on to focus on the awareness of MLCS, leading to barriers and challenges to the adoption rate of MLCS application within the SME sector in developed countries, and how advanced technologies could offer a more comprehensive model towards maintaining the security for zero-day attacks. The paper takes a closer look at the barriers to adoption of MLCS within the UK SME market and, based on the survey conducted within this paper, asks the question of why this adoption rate is low. As shown in the flow diagram above, Section 2 of the literature review puts a focus on discussions on and understanding of ML and its concepts. Section 2.1 sheds light on ML and its three subcategories: supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). Section 2.2 shares success stories on technology giants and their application towards MLCS and ML applications in general, as a way forward. Here, the different types of methods are showcased and lend a hand to the techniques and algorithms distributed to obtain the best effective solutions for MLCS already applied in the market. Section 2.3 goes on to explore the relationship and views between SMEs and their understanding of cybersecurity, especially in light of the recent pandemic and having to work from home. Vulnerabilities are discussed and digested to give a fair view of the current climate of cybersecurity. Section 2.4 further explores the SMEs' view of machine learning and shares case studies on different scenarios of SMEs involved in MLCS. This section pulls in the roles and responsibilities of government bodies and how policy and government involvement become paramount in yet again paving the path to a clear MLCS application.

Section 2.5 takes the journey of facing the challenges of cybersecurity for SMEs and compares varying literatures and how different developed countries under different jurisdictions have helped understand some of these challenges. Past experiments on the vulnerabilities of technology are also explored, leading back to how organizations are able to handle and overcome these barriers and challenges. Section 3 goes on to further explore, through methodology, the questionnaire survey targeted and focused on for this study in order to understand the awareness of MLCS of UK SMEs. The findings and analysis are discussed in Section 4, leading to the final Section 5 of outcomes and conclusions.

## 2. Literature Review

Within developed countries, understanding how SMEs view cybersecurity and machine learning is important. This then provides a good landscape to template against for those countries further developing in their technology to combat cybercrimes. Section 2.1 gives a brief introduction to the understanding and origins of machine learning. This section gives importance to the ML roots of AI. The section goes on to explore how ML is broken down into categories in order to help with managing Big Data and how using public datasets could help in future experimentations for testing and using MLCS technology. In order for MLCS to succeed, the labelling of data is important within these datasets in being able to analyze and obtain the right information processed from it. Section 2.2 covers the success stories of application of MLCS in industry. This is particularly important, as having success stories such as these and examples of how larger companies are using MLCS give a positive impact on how SMEs can use these templates to help protect and secure their data in the process. Whilst they may not be direct cut and paste applications, they are examples nonetheless of what works and what is still a learning curve for MLCS applications. These success stories show SMEs that with correct application, MLCS can work to their advantage and help to prove the method that could benefit the SME ecosystem. This section also shows that MLCS methods are worth adopting and worth exploring. Raising SMEs' awareness naturally increases the adoption rate of MLCS within the SME sector of developed nations, in particular the UK. Section 2.3 discusses the recent changes to how SMEs have changed in the way they work due to the pandemic, highlighting the issues raised for cyber security and how it is now becoming an important subject to talk about regardless of the industry the SME is in. Section 2.4 goes on to look at SMEs' understanding of machine learning through various examples experienced within the UK and its cyber agencies, including examples from developed countries. Section 2.5 completes the literature review in taking a close look at the challenges of cybersecurity for SMEs within the UK SME market and comparing those to other developed nations inclusively.

### *2.1. Machine Learning–Understanding and Origins*

In 1968, Arthur C. Clarke imagined that by the year 2001, a machine would exist with an intelligence that matched or exceeded the capability of human beings. By the 1980s, the film Robocop encapsulated AI technology through its robotic creation using ML and its algorithms [5]. AI and ML capabilities go far beyond the expectations of conquering human hobbies but lend further into everyday events in our daily lives. Professor Stephen Hawking, a world-renowned scientist, in an interview with the BBC in 2017 [6], discussed how efforts had been made to create thinking machines that potentially could pose a threat to our very existence. Hawking added that,

"The development of full artificial intelligence could spell the end of the human race."

Machine learning (ML) on its own stems from a branch of AI and is defined by computers being able to develop a model and learn over time without prior learning and then improve this model like a human [7]. Over time, the computer starts to develop and improve based on its interactions, as its software grows and develops. In a paper by Hewage, C. et al. (2018), one example of AI usage was to model polyalphabetic ciphers for decryption, in other words, to break the code following a set of sequential mathematical calculations and models of evaluation. Hewage's paper discussed select traditional algorithms such as hill climbing and genetic algorithm and simulated annealing to decrypt sample codes [8]. Similar to its predecessors and founders in code breaking back in 1941, the Enigma enciphering machine, which was used by the German army to send messages securely, was later on succeeded in its code breaking by the famous Alan Turing, who played a key role in his invention of the machine known as the Bombe, which significantly reduced the work of the codebreakers [9].

In the case of cryptology, designs of such algorithms in ML fundamentally lay within the strong structures of cryptology. As cited in the paper by Hewage [8], "Cryptology is the art and science of making and breaking 'secret codes'". Hewage's paper goes on to divide

cryptology into the two sub-divisions of cryptography and cryptoanalysis. Cryptography is the transforming and securing of the original data, whereas cryptoanalysis analyses the data to decrypt its encryption. Hewage's paper focuses its decoding using algorithms that were inspired by nature to tackle complex problems, in particular, looking at the ant colony optimization and how social behaviors influence the findings of the shortest paths leading to the end goal [10]. The algorithms used were hill climbing to decrypt sample codes as ways and means to obtain results through collection of data and getting results at each step of the "climb", whilst the genetic algorithm took an evolutionary approach. The results mutated and changed over a time period. Another algorithm discussed was simulated annealing, which was a process of heating and cooling and potentially trying to reach a local maximum to gain results. Worse solutions were discarded, to obtain the best possible solutions available. The paper unites the understanding of AI and furthers its categories inspired by nature.

ML can be divided into three subcategories of supervised learning (SL) (task driven), unsupervised learning (USL) (data driven), and reinforcement learning (RL) (learning from errors). In order to understand the advantages and disadvantages of how these algorithms work, a dataset is always used and injected into the algorithms. These datasets are then classified as labelled and unlabeled data [11]. ML is unable to move forward unless there is a dataset to work with. According to Buczak, A.L. (2015) [12], there exists a variety of datasets to choose from depending on the experiments being conducted. For the interest of ML algorithms, the public dataset was discussed. DARPA 1999 and KDD 1999 are amongst many datasets that have been used in the past and continue to be used in the public domain. These datasets that now contain more than 4 million records are difficult to maintain and require human intervention when it comes to labelling the records. How these datasets are labelled will define the type of category of ML utilized to move an algorithm design forward. These datasets sit very nicely under the DPA (Data Protection Act) 2018 and UK GDPR (General Data Protection Regulation) within a developed nation such as the UK. DPA and GDPR are important policy instruments regulating the framework for cyber security as well as data protection. This is important in the data mining and uses of datasets when experimenting with ML [13].

In ML, the first category is SL and is driven by tasks. It refers to the most basic types of ML, where the learning algorithm is developed on data [12]. Buczak further explains that SL can be further categorized into classification and regression. Classification refers to data points being set. Examples of classifications in real life include predictive text in tweets in Twitter and product reviews in Amazon and eBay. Algorithms used here are support vector machines (SVM) and naïve Bayes (Bayesian). Regression is used to predict continuous values, and examples of the algorithms are decision trees and neural networks. Real life examples include improving healthcare [14], calculating temperature, insurance premiums, pricing, and number of workers to the revenue of a business.

The second category in ML is USL. This uses datasets that are unlabeled, which means that human labor is not required to make the dataset machine-readable, thus allowing much larger datasets to be worked on by the program [13]. USL has two categories, namely dimensionality reduction and clustering, using many algorithms such as decision tree, random forest, missing values, principal component analysis (PCA), neural networks, fuzzy logic, and Gaussian. Dimension reduction focuses on data compression, and hence reduces storage space, leading to reduced computation time, and helps remove redundant features. Clustering refers to the task of dividing data into groups. Real life examples include identifying fake news, implementation of a spam filter, identifying fraudulent or criminal activity online, and marketing campaigns.

The third category in ML is RL, based on the psychological concept of conditioning. RL here works by putting the algorithm within a working environment with an interpreter and a reward system. The output result is then decided by the interpreter whether it is favorable or not. RL enables interactions with an environment through the means of a

machine. An example of this is repeatedly playing a video game, providing a reward system when the algorithm takes an action. AlphaGo, the online game is an example of RL.

The next section shows how MLCS has proven to work well in big technology companies and how the uses of ML technology and its methods have extracted success stories for SMEs to learn from and perhaps even apply at their level in industry.

### 2.2. Success Stories of Machine Learning Cybersecurity in Big Technology Companies

The information in Table 1 was collated as success stories of ML techniques used by big technology companies using ML methods, techniques, and algorithms. It also shares reference points and more success stories of where MLCS has benefited and helped these companies in securing their own internal systems from cyber threats. Reference to the legend is required to further explain the abbreviations in the table below.

**Table 1.** Success stories of MLCS techniques in Big Giant Technology Companies.

| Company | ML Method | ML Techniques | Ref. | Outcome and Results |
|---|---|---|---|---|
| Siemens Cyber Defense Centre using Amazon AWS | SL and USL | C, R, DR, and CLR | [15] | Build an AI-enabled, high-speed, fully automated, and highly scalable platform to evaluate 60,000 potentially critical threats per second. |
| PayPal, Visa, Mastercard | RL | SM and OL | [16] | Used machine learning in fraud management solutions to combat payment fraud. Using static models to identify fraud at a given moment by sifting through millions of past transactions. Identifying historical patterns of fraud and on self-learning techniques to adapt and recognize evolving fraud patterns |
| Darktrace in NHS | USL and RL | C, R, and OE | [17] | Uses machine learning to monitor raw data, such as cloud service interactions, transferred on a network in real time, without disturbing business operations and transactions. It also provides a direct view to all digital activities by reporting ongoing attacks or anomalies |
| Google—Gmail | SL | C, R | [18] | Used machine learning via filtering not just incoming spam but identifying other abuses like Denial-of-Service (DoS), virus delivery, and other imaginative attacks. |
| Tesla | USL | DR | [19] | Used machine learning to secure Wi-Fi and browser vulnerabilities using 0-day exploits to limit tampering with autonomous vehicles which can be disruptive |
| Facebook, Twitter, Myspace | SL | C | [20] | Developed machine-based classifiers to recognize precision in social spammers |

Legend: supervised learning (SL), unsupervised learning (USL), classification (C), dimensionality reduction (DR), regression (R), reinforced learning (RL), own experience (OE), static models (SM), own learning (OL), clustering (CLR).

Table 1 above references recent articles on the internet written by a variety of technology magazines, suggesting that Amazon's AWS (Amazon Web Services), Google's Gmail, and Facebook are all using their ML knowledge towards their cyber security models to advance their threat detection. Stephen Schmidt, Amazon Chief Information Security Officer (CISO), mentioned that Amazon had a duty of care to ensure the online safety of millions of people across the world, leading back to their cyber security structure. Siemens Cyber Defense Centre, which uses Amazon's AWS, went on to build an AI-enabled, high-speed, fully automated, and highly scalable platform to evaluate 60,000 potentially critical threats per second. This success story has then subsequently improved their cyber security and its threats reduction. In Table 1, it is also highlighted that Amazon used ML algorithms such as decision tree in its AWS Services and has expanded its services through Amazon's Macie on which its design was to embed its intelligence to protect the network and works of SL and USL methods [21].

In another article posted on CSO online [18], in order to analyze threat endpoints on mobile devices running on Androids, Google was able to use ML in identifying and removing malware from these devices. As clearly shown in Table 1 above, Google mail (Gmail) has seen success stories in its spam filtering, not just incoming spam but by the use of machine learning in identifying other abuses, such as Denial-of-Service (DoS), virus delivery, and other imaginative attacks [20]. Based on these ML methods, Amazon launched a new service to classify its data storage under the SL techniques of ML.

Table 1 above also shows the applications of a UK cyber security start-up company, Darktrace, a company that had seen success around its ML solutions since 2013 [22,23]. Darktrace used algorithms within its software package to spot attacks within one NHS agency's network, and the threat was then mitigated without causing any damage to that organization. When WannaCry was the top cyber threat back in 2018, all Darktrace customers were not harmed, as the ML algorithms were clever enough to intervene and create a safe environment for them [19]. According to Vähäkainu, P. (2019) [18], Darktrace uses its own mathematical algorithm, Enterprise Immune System (EIS) technology, and utilizes this ML technique combined with the Bayesian algorithms and other mathematical principles in order to detect anomalies for cyber threat detection within a network. Vähäkainu describes the technology using Bayesian probability theory and how Darktrace monitors raw data, such as cloud service interactions. Vähäkainu also explains how this data was then transferred onto a network in real time, without disturbing business operations and transactions.

Other companies to take up MLCS are companies such as PayPal, Visa, and Mastercard. These companies use deep learning algorithms to identify and prevent fraudulent behavior within milliseconds before, during, and after a transaction, as reported in the article written in November 2020. Mastercard also had experienced over 200 fraud attempts per minute, which allowed them to also utilize the ML algorithms to combat cyber security threats. Mastercard too chose to implement deep learning algorithms within their network.

Another article in the *MIT Technology Review*, dated April 2020 [24], explained how hackers were trying to trick Tesla's program into veering into the wrong lane whilst driving. However, Elon Musk's investment in ML showed strength in trying to overcome this issue. Table 1 goes on to show that in a similar study [18], Tesla used ML to secure Wi-Fi and browser vulnerabilities using zero-day exploits to limit tampering with autonomous vehicles, which can be disruptive.

Various other research lends particular interest to MLCS in action and how particular use of ML algorithms specifically enhances the interest of the applications used. In particular, e-commerce applications provide an added advantage to customers to buy products with added suggestions in the form of reviews, similar to the design of the likes of Amazon and eBay. In a paper by Uppal, S. (2019), the author gave importance to how these reviews become useful and form impact for customer engagement on wanting to purchase products. However, whilst most reviews are positive, many can create problems if they are less savory in nature and if customers not being able to segregate useful ones from those that are nonsense. Uppal's paper pays attention to the need for an approach which will showcase only relevant reviews for the customer's interest. Uppal's paper suggests the "Pairwise Review" relevance ranking method, which is based on their relevance of the product and avoids showing irrelevant reviews. ML algorithms used here were SVM, random forest, neural network, and logistic regression, being applied to validate ranking accuracy. Out of all four applied classification models, random forest gave the best result and achieved 99.76% classification accuracy and 99.56% ranking accuracy for a complete dataset using random forest. This success story showed that ML usage is becoming more applicable in its design for everyday application as well as cyber security for protecting the network, in this case protecting the integrity of a sound business with a genuine reputation [25].

In real life applications of the previous section of ML techniques and its algorithms, technology giants such as Amazon, Google, and Facebook all have been gradually ramping up their security models in using AI and its usage of ML. These technology giants have

used ML to focus on how they can use the technology to improve their customer service experience and further develop their customer engagement and behaviors and complement their cyber security. These technology giants have also created ML products to protect their own customers from cyber threats [26].

In the same paper by K. Lee et al. [15], it was observed that malicious spammers would exploit social media systems of these technology giants such as phishing attacks, malware, and promoting affiliate websites, thus leading to the development of detecting spammers in social network companies such as Twitter, Facebook, and My Space. Developing specific classifications techniques enables the detection of email spam and phishing approaches that rely on data compression algorithms, machine learning, and statistics that could inform the further refinement of many proposed approaches. Lee's paper uses SL techniques based on support vector machine (SVM) with its high precision as well as low false positive rate with its information and data feeding into the SVM classifiers.

ML's celebrity status is covering nearly all disciplines including that of sports analytics in visualizing impact to assist in decision-making in making sports performance at its peak, as explained in Jayal, A. et al.'s paper. Jayal's research uses big data approaches and analysis of approach-based structures in integrating problem-based learning through interactive visualization, simulation and modeling, geospatial data analysis, and ML, amongst various other big data techniques, in particular ML and its algorithm usage in sports. The approaches of clustering techniques, survival analysis, artificial intelligence, rule-based approaches, graph-based approaches, and inductive logic programming plus neural networks and deep learning allow for greater understand of identifying a general-purpose toolkit that can be used with the help of data reduction and data mining and analytics approaches in sports [27].

Through these success stories, as shown from the diagram above, there is certainly an overlap in the types of methods being used, and this lends a hand to the techniques plus algorithms distributed to obtain the best effective solutions in the market to combat cyber threats through various cyber security software packages. The next section will lead on to the SME's view of cyber security and UK SME's adaptation of MLCS.

### *2.3. SME's View of Cyber Security and Adaptation of MLCS*

At the start of 2020 there were 5.94 million small businesses (with 0 to 49 employees) in the UK, accounting for 99.3% of the total business as recently reported by the National Federation of Self Employed & Small Businesses (FSB) [28]. The same set of statistics has shown that UK SMEs account for 99.9% of the business population equivalent to 6.0 million businesses. According to the definition by the UK government, micro-SMEs hold less than 10 employees and an annual turnover under €2 million, small SMEs have less than 50 employees and an annual turnover under €10 million, and medium-sized SMEs have less than 250 employees and an annual turnover under €50 million. Between 2019 and 2020, the total business population grew by 113,000 (1.9%). The COVID-19 pandemic has caused the UK to face challenges effecting the economy, and SMEs alongside other organizations have made a shift from physical shop windows to virtualizations in cyberspace [29].

According to the Office of National Statistics (ONS) reporting on December 2020, temporary closures, a shift to online shopping, and reduced travel meant the first wave of the coronavirus (COVID-19) had an enormous impact on business, and some industries felt the impact far worse than others [30]. Whilst some industries shrank by up to 90% in April and May, others recorded some growth. In particular, online shopping grew far more than its pre-pandemic trend, and our cyber footprint exploded, seemingly having no boundaries [31].

For SMEs to further reach their network and grow their businesses, online activities have seen a massive rise in how SMEs have had to change the way they worked to accommodate this change. SMEs have had to change their technology and organization, but most importantly change how they work with their staff, with working from home to making sure their business data are kept safe and secure. Whilst larger organizations

have had the benefit of many departments cushioning various corners of the business with the right people being paid the right money to support the organizations, this scenario is not the same for SMEs. With a smaller group of people to manage the business and controlling the growth rate, SMEs fall into a niche category of experts that potentially have to understand and know everything about the business and be flexible in how work is conducted and administered.

In light of these challenges and changes to SMEs and hybrid working conditions, an exceptional rise has been seen on the usage of Internet of Things (IoTs). Employees working from home are having to juggle personal and business life through using personal devices to access business data [32]. Data shared with each other and the need to share data in particular ways have now become important in recent events of needing to work from home during the current pandemic of 2020. The pandemic has brought to light the need for using IoTs, such as daily usage of phones, iPads, and other smart devices. These IoTs are being used in industry to keep up with the growing trends of getting information faster, whilst having advantages to the ever-growing IoTs in these industries and devices talking to each other in a connection of networks across cyberspace, which allows for transfer of data to happen quickly and efficiently. This in particular is advantageous to the SME industry for its size and its ability to be flexible in how their employees work and the changing lifestyle in which SMEs need to grow.

However, this scenario within the functions of an SME is now presenting numerous challenges, including those related to privacy, security, and data breaches, or those pertaining to ethical, legal, and jurisdictional matters. IoTs cover a broad range of proprietary hardware and software that often use different data formats, networks, or communication protocols, and physical interfaces resulting in technical challenges. MLCS methodologies allow for the analysis of SME business and arise to management questions on how multiple interactions and complexities arrive from being connected to the internet. These large quantities of data are often private and sensitive, transferring data along the way. Disadvantageously, this creates a wider security attack surface for potential malicious activities to occur.

Looking at how IoTs and ML have clearly moved forward positively and making it easier to manage, humans now cannot even imagine life without technology. Hard as it is to imagine, the realization has taken one step further in that the pandemic of COVID-19 in 2020 has accelerated the usage of IoTs and its applications of MLCS into new realms humans perhaps cannot even understand. Even the likes of Chatbots have emerged to manage online interactions linked to the use of AI applications. Chatbots [33] have replaced people online, and ML is now learning everything about us and how humans behave. ML in its integration into IoTs is now evolving in how we interact online and adapt to our needs and surroundings. The desire for humans to interact with machines is vital. It is no wonder that the assumption of people's awareness is at stake.

### *2.4. SME's View and Support on Machine Learning*

The UK's answer to providing intelligence and information assurance to the government and armed forces is the Government Communications Headquarters (GCHQ) [34]. The GCHQ is an intelligence and security organization with a mission to keep the UK safe.

The National Cyber Security Centre (NCSC), under the parent body of GCHQ and other national security centers, offers online guidelines for SMEs and business on how they can avoid cyber-attacks. Following these guidelines helps SMEs give awareness and shape their business to keep their data as safe as can be. The guidelines follow a set of rules such as backing up SMEs data, protecting organizations from malware, keeping IoTs safe, using good structured passwords and management to protect the data, and how to avoid phishing attacks, amongst many other tips and tricks. Most of these guidelines give helpful hints and share knowledge on how to develop a state of awareness and be diligent in keeping information safe [35].

In recent news published September 2020 by GCHQ, ten tech cyber security start-up companies using AI, Data Science, and ML were selected to benefit from the 12-week support program, based out of GCHQ's Manchester office. These included firms which use AI to alert haulage companies to stowaways in their containers, data to determine how busy trains are to manage social distancing, and how AI and ML were used to identify and prevent the spread of fake news [36]. In April 2019, guidance was being written by the National Cyber Security Centre website (NCSC), which is now part of the GCHQ, that offered information on assessing intelligent tools for cyber security in the form of AI and ML. The NCSC provides a single point of contact for SMEs, larger organizations, government agencies, the general public, and departments, and also collaborates with law enforcement, defense, the UK's intelligence and security agencies, and international partners.

These methods adopt the stranger danger policy in helping SMEs move forward. Whilst this is useful, many SMEs fall short due to how they go about securing their data rather than getting their hands dirty for prevention.

SMEs, due to their structure and economic characteristics, can be extremely damaged when a cyber-attack takes place. In a 2020 study by López, M.Á. on intelligent detection, the author outlined the different scenarios of cybercrime and what can be done to compensate the situation [37]. Here Lopez proposed an intelligent cybersecurity platform, which had been designed with the objective of helping SMEs to make their systems and network more secure and robust. The proposed aim of this platform was to provide a solution optimizing detection and recovery from attacks. The proposal applies a proactive security technique in combination with both machine learning (ML) and blockchain. The proposal, which is part of a funded project by the Innovation and Development Agency of another developed nation country, Andalusia, Granada, Spain (IDEA) (IASEC project), allows for the provision of security in each of the phases of an attack in helping SMEs in prevention, avoiding systems and networks from being attacked. For SMEs, using various different software to manage their security information and event management systems (SIEMs) is very important in helping organizations become compliant and to have the infrastructure in place to help with any breaches. Lopez et al. proposed providing resources to optimize detection and self-recovery of systems and services after suffering an attack, creating a solution to allow detecting, and dealing with fake publications on the Internet, protecting IoT devices and Industry 4.0 from the most relevant attacks for SMEs, and detecting and avoiding fake news and hoax spreading. These objectives are tackled by combining both smart systems and blockchain. Blockchain here in the proposal is used to improve the security systems by protecting data integrity in a secure and transparent way.

Another study by Rawindaran, N. et al. (2021) [38] explored how early detection of cyber-attacks is important through SIEMs, especially in the cycle of network security. Intrusion detection and prevention systems (IDPS) were experimented with, and commercial network intrusion detection systems (NIDS) versus open-source devices were compared to combat cyber-attacks. These IDPS devices all came with their own SIEMs to track events and send alerts to become part of the cycle of IDPS. Amongst those that were discussed were SolarWinds, Cisco, Tripwire, Wireshark, and Splunk, to name a few. Protection of data, as evaluated and discussed in Rawindaran's paper, is the reason why IDPS systems have come into force more within the SME market [38].

Both Rawindaran, N. et al. and Lopez, M.A. et al. agree that IDSs can be network-based (NIDS) and host-based (HIDS) and can monitor and analyze network traffic in real time together with analyzing records, databases, and other elements in a host to detect possible intrusions. IDS can also be grouped according to the type of detection technique, being signature-based and anomaly-based [38].

ML techniques and algorithms have now contributed largely to how data can be classified, labelled, and ultimately managed under the umbrella of AI. The ability to use techniques such as supervised and unsupervised learning has helped in getting Big Data within this cyber space, through various classification, regression, and clustering activities.

These activities allow for outcomes to be predicted. ML mathematical algorithms all compound to how data is treated and managed to produce the outcomes and predictability required to contribute to economic growth in societies moving forward.

In Lopez, M.A. et al.'s [37] proposal, ML techniques were used for data collecting, testing, and evaluation, and the main goal was to determine the most efficient algorithm for intrusion detection. ML algorithms for supervised detection were compared, such as C4.5 (decision tree), Bayesian network, random forest, support vector machines (SVM), and artificial neural network (ANN). The study performed measurements from different sampling data, and the results showed that C4.5 was the most precise among the studied algorithms. Finally, another proposal was to build a solution focused on cyber security for a smart-home or smart-office, applying two variants of long short-term memory (LSTM), which is a type of neural network.

SMEs are all aware of Denial-of-Service (DoS) and Distributed DoS (DDoS), malware, or web-based attacks, as they are some of the most common security incidents around. Lopez explains that when a server suffers a DoS attack, the system records in the smart contract those IP addresses that are involved in the attack, creating new blocks every 14 s through block chain technology. Each user in this network now has an updated list with malicious addresses in the interval, allowing the security people to take actions for attack mitigation. This solution can be extended to DDoS attacks using a dataset that has been accurately obtained using the random forest ML algorithm for model building. Similarly, for structured query language (SQLi) attacks detection, datasets are applied ML algorithms such as decision stump, naïve Bayes, Bayesian network, and radial basis function (RBF) network, which is an ANN. The most efficient algorithm was decision stump. Naïve Bayes was then used to classify SQL queries as malicious or legitimate. Both grammar and SQL syntax were taken into account and extracting features from language and defining rules. Training several classifiers, such as SVM, ensemble bagged trees, or ensemble boosted trees was important, and it was identified that in this case, the best result obtained was the decision tree model. Another attack that is common to SMEs is the domain generation algorithm (DGA), and this can be detected by analyzing DNS traffic in pseudo-real time. DGA is used to generate new domain names and IP addresses for malware's command and control servers. Here the proposal enables filters and non-resolved DNS requests and identifies those hosts showing the highest peaks for this value for detection. From this study of Lopez, it is very apparent that the three most common attacks to SME infrastructure can be identified and prevented by the correct use of ML coupled with block chain technology to protect the business. By using the right SIEMs together with the IDPS and NIDS/HIDS, SMEs can be educated in the right direction to be able to make the informed decisions they need to make.

Another study showed the evaluation of ML algorithms for anomaly detection is performed through the ALICE high performance computing facility at the University of Leicester [39]. The impressive computer had 64 GB of RAM, two Ivy Bridge CPUs at 2.50 GHz (20 cores in total), and 2 × Nvidia Tesla P100 GPU cards. Python 3.6.8 was used to run the service on an Enterprise Operating System3 (CentOS Linux 7). The classical ML algorithms were implemented using the Scikit-learn 0.21.3 ML library. The deep learning algorithms were implemented using Keras 2.3.04 neural-network library on top of TensorFlow 1.9.05 to enable the use of GPU. Sigmoid and SoftMax functions were also used for binary and multi-class classification. Pandas6 and NumPy7 library packages were used to manipulate and analyze the raw data. This evaluation looks at a comprehensive analysis of the ML algorithms, with the result being that the random forest (RF) algorithm achieved the best performance in terms of accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curves on all datasets given. The main contributions of this paper were that the currently available datasets containing the most up-to-date attack scenarios were used, and ML anomalous detection was applied. Binary classification and multi-classification based on the performance metrics were used and produced the best-fit algorithms for the anomaly detection challenge. This same study shared the research

community's and the SME cybersecurity industry's insightful knowledge and suggestions regarding suitable ML algorithms to support cybersecurity.

In all its glory and complex structure, ML is playing an important part in the way we handle attacks for cyber security and protecting our data. In a paper presented by Gupta, A. (2021) [40], the authors gave clarity to the various applications of ML in cyber security within the SME market. The ability for ML to detect malicious events and prevent attacks are the top reasons to use ML within the cyber security infrastructure and start using devices and technology than can support anomaly detection for zero-day vulnerabilities and protection of networks, endpoints coupled with application security, and user behavior. ML usage in IoT comes in second as the incorporation into mobile gadgets such as Google and Apple's Siri have become important in the cyber security ecosystem. Various other uses of ML will go on to include human analysis and make our jobs easier in terms of being able to filter data, review millions of login details, pass information on to human analysts, and minimize notification and build a complete AI system to support the system.

The next section looks at how SMEs' views and support of MLCS lead to barriers and challenges within the industry and how this can be overcome.

### 2.5. SME's Cybersecurity Barriers and Challenges

Whilst there is a huge advantage to using ML in the SME industry, the disadvantages include dataset availability for testing, and the fact that information can be mixed up, as well as the need for information to still require ground truthing, according to Gupta [40], as human intervention in creating the mathematics and the models is still unfortunately required and the margin of human error is still to be defined. The degree of human intervention is still strong. Lopez, M.A. et al. [37] highlighted barriers such as resources, and not having enough knowledge to set up efficient security systems, such as SIEMs, to challenges in implementing a security platform that provides this knowledge through means of ML techniques. Whilst the architecture is scalable, SMEs rely on micro-services for detection and recovery when an attack is predicted to occur.

SMEs have become most vulnerable to cyber-attacks due to their unique ecosystem. One reason could be due to the potential shortage of cybersecurity knowledge and resources that exist in the SME organizational structure. SMEs have become put into positions of exploitation, whereby the likelihood of cyber-attacks come at a high price in experiencing cyber incidents. In a recent paper by van Haastrecht, M. et al. (2021) [41], SMEs struggle to cope with the rise in cyber security threats leading to intuitive, threat-based cyber security risk assessment approaches for the least digitally mature SMEs, using a socio-technical cyber security framework to help contribute towards the needs of SMEs. The works of van Haastretcht use both a framework and the ADKAR (awareness, desire, knowledge, ability, reinforcement) change management model of Hiatt [42] to guide the research in covering the social dimensions needed to be considered in SMEs. Coupled with five main aggregation strategy classes applied, such as weighted linear combinations, weighted products, weighted maxima, weighted complementary products, and the Bayesian network, the results are able to determine if the application within the SME was too simplistic or needed advanced care. The framework was then applied on SMEs that were divided into further four categories, as suggested by the European DIGITAL SME Alliance [43]:

- start-ups,
- digitally dependent SMEs,
- digitally based SMEs, and,
- digital enablers.

In summary, digitally based SMEs and digital enablers were advised to use a more comprehensive risk assessment approach and maturity model due to the expertise available within the SME organization to cope with building trust in cyber security along with standards and policies in place. Digital enablers were also prime candidates for using more advanced aggregation strategies, such as Bayesian networks, due to having the cyber security expertise and data required to make these solutions successful. For start-ups and

digitally dependent SMEs, threat-based risk assessment approaches worked better based on non-aggregated or intuitive strategies by focusing on the real-life threat environment to accommodate feelings of competence and relatedness by ensuring optimal organization and employee motivation and doing what is right. Van Haastrecht, M. et al. goes on to explain that one size does not fits all and the type of SMEs matters, and the intellectual knowledge contributes to the success of its cyber security landscape. The barriers here reflect that SMEs cannot adopt a "cut and paste" style of understanding cyber security and its threats like how larger organization can.

In another article by Tam, T. et al. (2021) [44], another developed country is examined through the lens of Australia. Lessons are shared of how developed nations such as Australia deal with their SMEs and how they are faced with cyber security challenges. Large organizations within Australia have always been early adopters of cyber security scenarios often having the workforce, finance, and environment to support the research and development in cyber threats. Tam explains that most cyber security lessons and conventions exists due to the result of early large-scale incidents such as NotPetya, Equifax, Wikileaks, etc., affecting mostly large organizations. Consequently, cyber security industry best practices, standards, and products are influenced by the needs of larger organizations. Tam also highlights that the technical landscape of an SME can potentially be very different from that of a large enterprise, making it impractical to apply solutions for the larger enterprises to smaller scale users. Taking an example back to the UK was the implementation of Cyber Essentials and GDPR. Larger organizations had an easier approach for implementation compared to that of SMEs purely due to their ability to be able to have the labor-power and the technical expertise to implement at a smoother rate. The small business IT technical architecture becomes another barrier to adopting a complete cyber security solution. Tam goes on to further explain that another major barrier for technical implementation is the need for a robust testing environment. Testing environments are achievable between larger organizations than SMEs in the context of this Australian example. Tam explores in this study that any cyber security solution designed to test a response to debilitating events requires a safe testing environment.

For example, denial of service (DoS) simulation tools can simulate a service overwhelmed with requests, resulting in legitimate requests not getting through. A DoS simulator, if implemented on a live system, would render the SME business IT infrastructure, e.g., website, unavailable to customers, or worse, jeopardize the overall system integrity and potential loss of business. Tam concludes that live environments cannot be used for stress-inducing tests. Consequently, businesses without a test environment will never be able to test the full suite of catastrophic scenarios as part of their incident response training. Tam goes on to discuss the importance of a test environment that requires substantial technical knowledge, time, and ongoing maintenance, which is only feasible in larger organizations and very rarely seen within the SME context.

In addition to barriers of technical challenges, Tam's study also highlights barriers such as human factors that contribute to SMEs having challenges in implementing the right cyber security choices, leading to organizational and process maturity of the SME sector. The complexity of implementing industry standards and having to bear the costs of cyber insurance, legal remediation, and costs of a data breach also contribute to why SMEs in Australia have found moving forward to protect their data sometimes impossible to keep up with. Tam's paper sits well with the given technology landscape that is very similar to countries such as the UK and hence will have similar cyber security concerns. similar. SMEs in Australia and UK hold similar societal profiles, thus sharing similar human struggles with cyber security. The conclusion to Tam's paper suggests that opportunities to apply non-traditional solutions to cyber security are becoming apparent through new found alliances, security paradigm, and the open source community for helping SMEs build up their defenses to combat cyber-attacks.

The literature review section above required the uses of various platforms in order to perform searches for the topic in concern for this article. The methodology is documented

in Appendix A of this article. The next section reveals the methodology applied in the survey questionnaire run in this paper to hear and listen to the voices of SMEs in the UK on how their impressions have been in these various cyber security topics, paying particular attention to the awareness and changes through the pandemic and how governments could make some changes in bridging the gap to a better and safer cyber landscape moving forward.

## 3. Methodology

This paper involved the distribution of a survey questionnaire to UK SMEs in which data were collected from multiple respondents. This survey questionnaire is filled and referenced under Appendix B in the Appendix section at the end of this article. The survey was distributed using the software Qualtrics [45], which sent this questionnaire survey to an extensive participation list of UK SMEs. The survey is one that can be used as a generic template in all developed countries; however, for the purpose of this research, the UK was chosen for this pilot study for its ease in participants responding back in a timely manner. The political and economic demographics of the UK are representative of other developed countries, thus providing a similar platform and landscape to the participation pool involved in this questionnaire survey. The participant pool acted as a benchmark for representing SMEs in developed countries, with the UK being the focal point of research. The figure below shows the participant selection criteria based on the population detected for this UK SME research.

The figure above shows the stratify method used in selecting the participation pool for the research. The method chosen was stratified sampling, which is a method of variance reduction. In a study by Acharya, A.S. (2013), data were divided into various sub-groups (strata) sharing common characteristics such as age, sex, race, income, education, and ethnicity [46]. A random sample was taken from each stratum. The advantages of stratified sampling are that it assures representation of all required groups in the population. The characteristics of each stratum can be estimated, and comparisons can be made. It also reduces variability from systematic sampling. According to Acharya, the limitations are that it requires accurate information on proportions of each stratum; furthermore, stratified lists are expensive to prepare. It ensures that at least one observation is picked from each of the strata.

According to the flowchart in Figure 2 below, the participants were selected from a variety of sources on social media sites and direct email contacts. Using Qualtrics, the survey was distributed across two groups. Group 1 included the social medial professional site LinkedIn, and group 2 covered contacts through social media news on Twitter. Further to this, a third group of participants was collected via email distribution. The survey was a time-based frame and given the timeline of two weeks. A timeframe is important for larger projects according to Greenfield (2002) [47]. This study was not an exception; as such, with a timeframe of two weeks, a condensed window was given for participants to fill in the questionnaire and for the study to receive quick and effective results. The three groups then formed the basis of results collected and samples from each group taken to be part of the survey results. The advantage of this method is to show fairness in the types of samples being targeted and the responses analyzed. It is also to make sure the demographics are covered across the SME landscape. The disadvantage to questionnaires is that the survey might miss out on more in-depth or abstract observations being recorded (Sarantakos, 2013 [48]).

The sampling strategy used to select participants was that of stratified sampling, as the relationships between different groups had to be observed (Kirby et al., 2000:339) [49]. Stratified selection guided by a timeframe result in a numerical number of participants filling in this survey and returning a completed filled survey. Using the stratify method, certain groups of SMEs were targeted to collect and complete the feedback, and this covered various ranges from different industries. Participant pools consisted of industries to include nurseries, healthcare, retail, technology, estate agents, amongst others, to fill

out this questionnaire. The participant feedback was anonymized unless the participant wished to be named and contacted for further research.

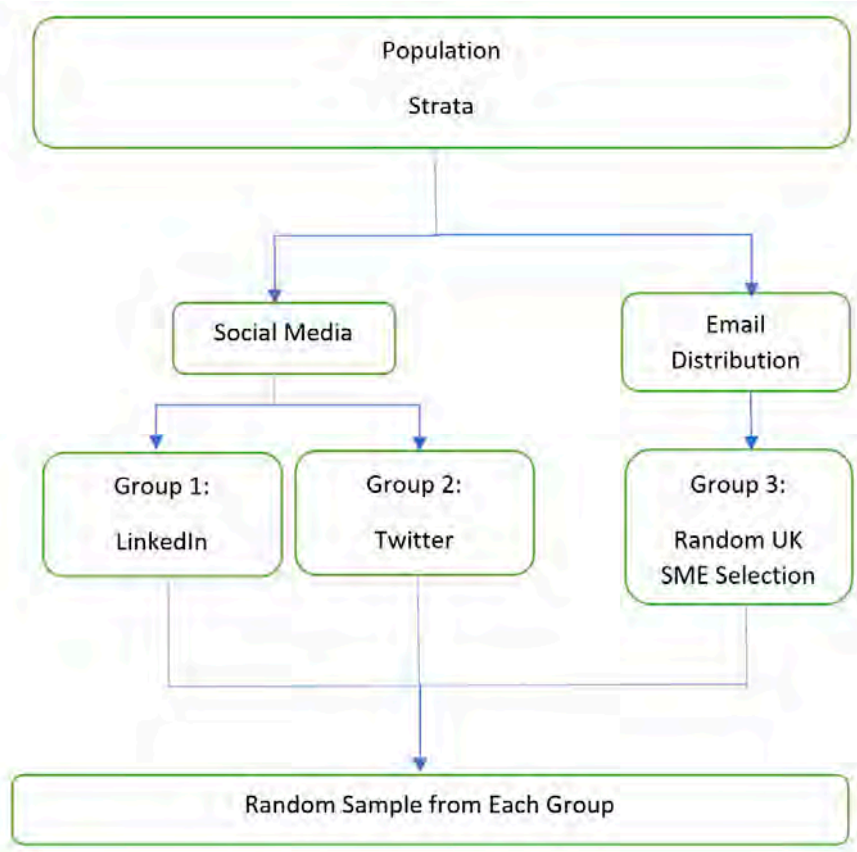


**Figure 2.** Stratified flowchart of methodology used.

Within Qualtrics, a new project was built, and twenty-one questions were filled in, a combination of multiple choice and text-based questions. Both quantitative and qualitative methods were used. The questions were sectioned into four sections:

(1) Details of participant and experience;
(2) Knowledge of cyber security and its packages;
(3) Knowledge of the machine algorithms used in the packages;
(4) Cost implications of machine learning in cyber security software packages.

It started by collecting information such as their role within the organization, as well as management, technical, and non-technical expertise as part of the stratify method. The questions focused on the individual's age range; identity; and their role, education, and industry. Next, the questions asked if the SMEs had any cyber security software packages in place to protect their business from cyber threats, and if so, to state the software, and if no, then to give a reason. This question then led to the detection of cyber-attacks and if machine learning was used as part of the options to secure their network and data. The survey touched upon costs of machine learning implementation and the satisfaction in trying to implement this solution with the current IT expertise and infrastructures in place. The survey came to an end by asking for opinions on how SMEs can raise awareness of machine learning, how this can be made better, and if it is appropriate to follow up on the responses moving further into the research in the future. The survey targeted different sizes of SMEs, micro, small, or medium, based on the number of people in the company. The SMEs will also get an opportunity to answer the relevant questions pertaining to their understanding and awareness of MLCS and their current cyber support packages. The ethics process, approved by the Cardiff School of Technologies Ethics Committee, was

followed throughout this study. Informed Consent was taken from all participants. No personal data were collected, and all the data were anonymized.

## 4. Results and Analysis

The awareness of machine learning cybersecurity (MLCS) within the environment of SMEs in the UK became the subject of this research, and the key questions and findings are discussed in this following section. Key questions that were shared were broken into four components:

- Details of participant and experience;
- Knowledge of cyber security and its packages;
- Knowledge of the machine algorithms used in the packages;
- Cost implications of machine learning in cyber security software packages.

These components formed the focus of the study, providing the basis of the discussion on which the SMEs' awareness was questioned. The broad view of the questions that were shared asked the SMEs what software they used for their cyber security package and if these software's had options for ML techniques. The questions also focused on SME participants' understandings of the configurations and algorithms used for these ML techniques and if they were being used within the SME sector. The questions also looked at the understanding of the costs of cyber security software packages with and without ML techniques. The next section provides the findings of the responses.

As part of the first component, the participant pool was carefully selected based on the targeted audience of UK SMEs using the stratify method, as outlined in the methodology section. The industries chosen included the following:

- Engineering, IT, and Consultancy;
- Healthcare;
- Hospitality and Service;
- Insurance;
- Other.

Other referred to a range of industries not covered under the main components above. These were research, distribution, garage services, property, printers, health and safety, estate agencies, retail, and logistics and supply chain. Figure 3 shows the breakdown of the industries as described above.


**Figure 3.** Industry participants who took part in the survey representing UK SMEs.

From the participant list surveyed, this paper also looked at the education level of the participants taking the survey. As seen in Figure 4, 26% of participants had a bachelor's degree as a base educational level.

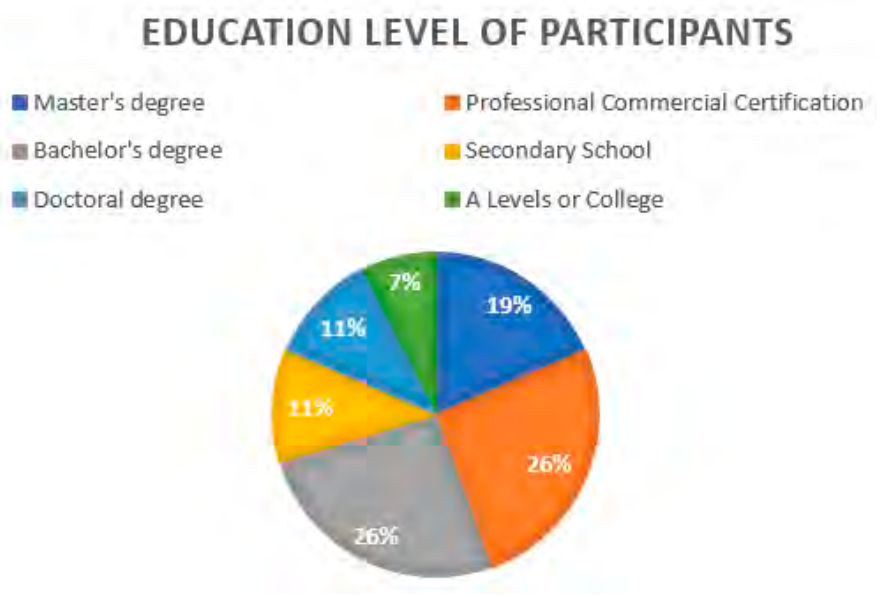


**Figure 4.** Education level of participants in %.

It was also noted that many participants held a university degree plus a professional commercial certification. Figure 5 below shows participants' positions in the UK SME.

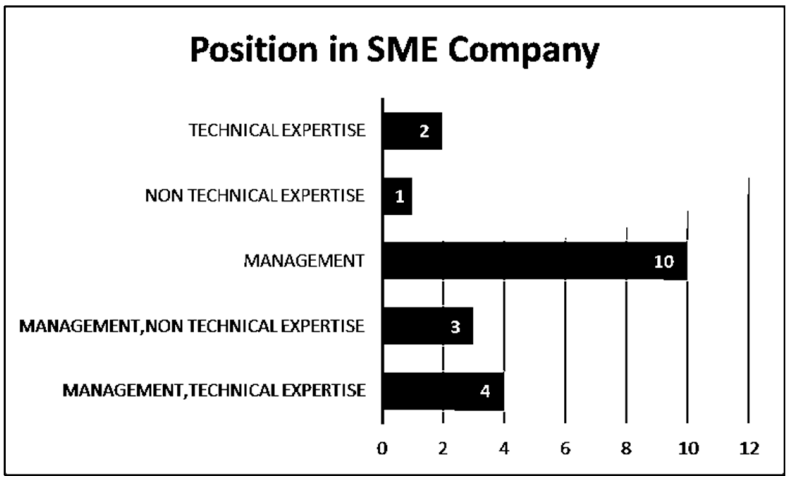


**Figure 5.** Position in SME company.

Figure 5 shows that the majority of the participants were from management and also had the technical knowledge to run their companies. There were a few selections that were in management that did not have this technical expertise. Others had exclusively technical expertise, and one was non-technical.

Figure 6 below shows two components reflecting age and identification of participants in the study. It was clear that there were more males in the field of this study responding than females, and the age range covered a higher proportion of participants aged between 36 and 55.

The second component revealed the findings based on the questions surveyed. Figure 7 shows the results of the question whether SMEs have cyber security software packages that protect their businesses from cyber threats.

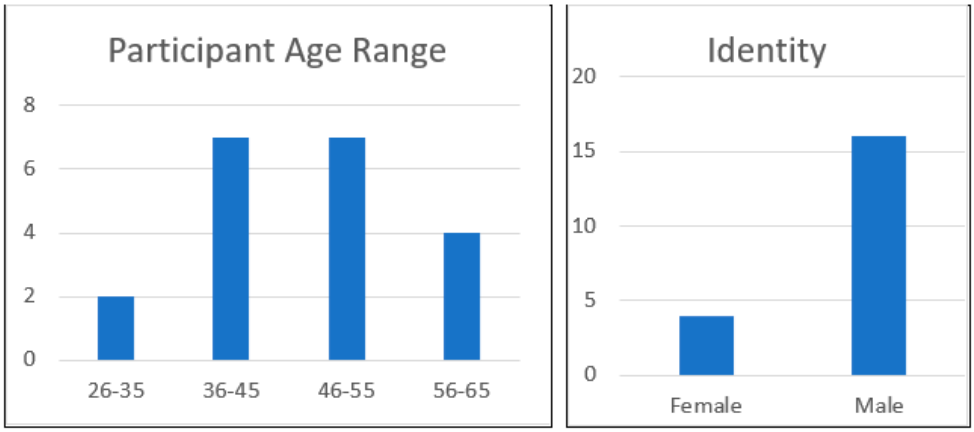


**Figure 6.** Two components reflecting age and identification of participants.

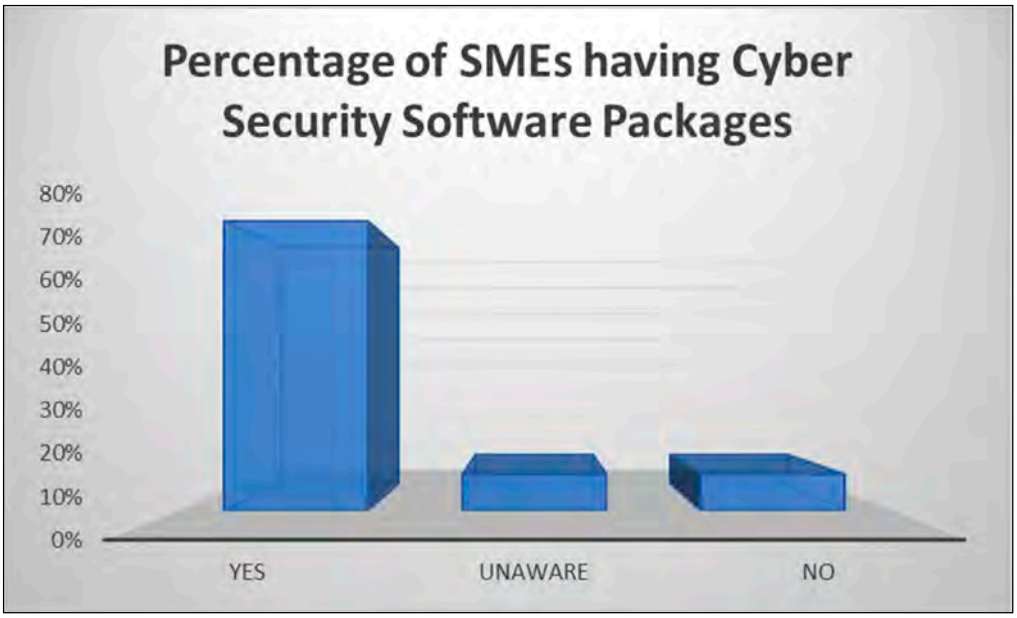


**Figure 7.** Percentage of SMEs having cyber security software packages.

The results indicate that 80% of SMEs have cyber security software packages in place, with a further 10% stating that they did not have these packages installed, and 10% stating that they were unaware, as shown in Figure 7.

Based on the above response, the packages identified are shown in Figure 8. Out of the 20 respondents that replied to the survey, a breakdown of all the cyber security software packages used is shown in Figure 8.

Based on the results shown in Figure 8, these software packages included Checkpoint Firewall, SolarWinds, Cisco, and Microsoft, amongst many others. Figure 9 shows SME participant awareness of the existence of ML in their cyber security software packages.

The third component showed that the proportion of SMEs showing awareness of the existence of ML in their cyber security software package was 30%. Those who said they did not know was 60%, and 10% were a definite "No" in their response to awareness of ML being embedded in their software packages.

Referring to the pool that said "Yes" in the 30% of those replied, we went on to drill down into the types of ML algorithms that they understood were present in the software packages based on the manuals and specifications provided with the package. In Figure 8 it can be seen that various types of algorithms were identified as being part of the ML that

existed within the cyber security software packages defined above in this paper. Amongst them were neural networks, Bayesian model, support vector, and deep network.

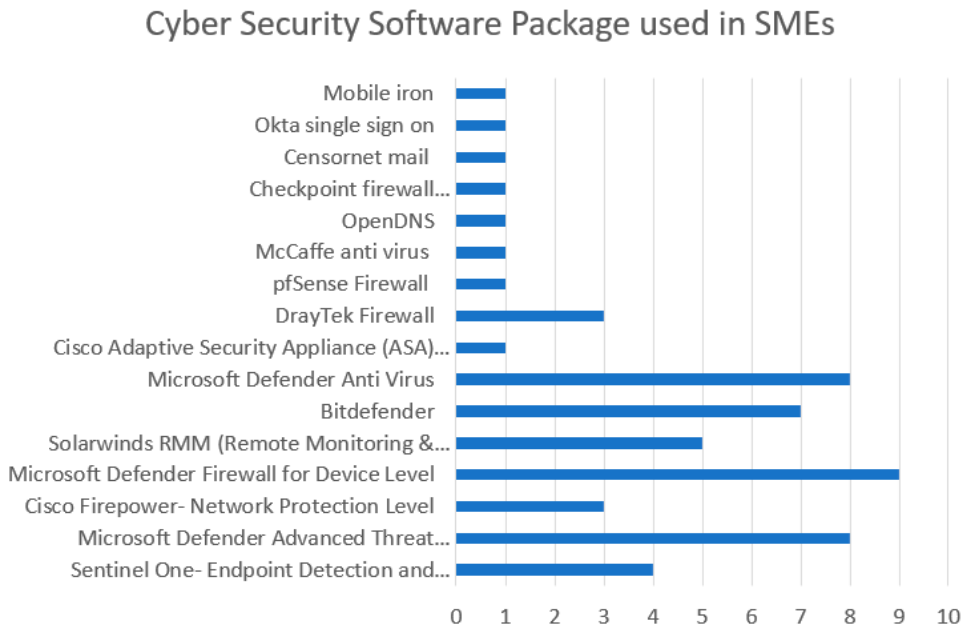


**Figure 8.** Cyber security packages used in SMEs.

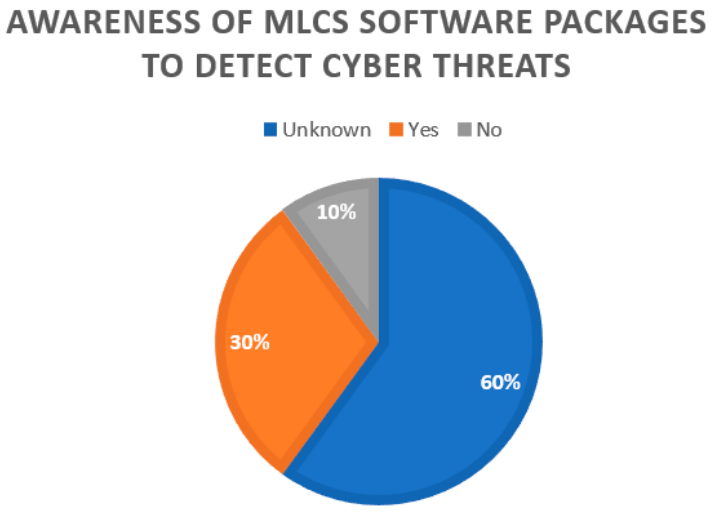


**Figure 9.** Awareness of ML in cyber security software packages to detect cyber-attacks.

As seen in Figure 10, there were a few participants from the pool of respondents that either did not know the algorithms or found they were not applicable to the software they were using.

Drilling down further, Figure 11 shows the actual algorithms that were being used rather than what was available as shown in Figure 10.

Figure 11 above represents the various algorithms that have been implemented in the use of ML based on what was purchased by the SMEs and their cyber security software packages. It was clear that in the majority of the cyber security software packages, neural networks and deep networks were used, Bayesian was not far behind, and Microsoft-owned Azure algorithms were a top interest in the growth of ML.

The final fourth component revealed the awareness of the price attached to ML in cyber security packages, as shown in Figure 12.

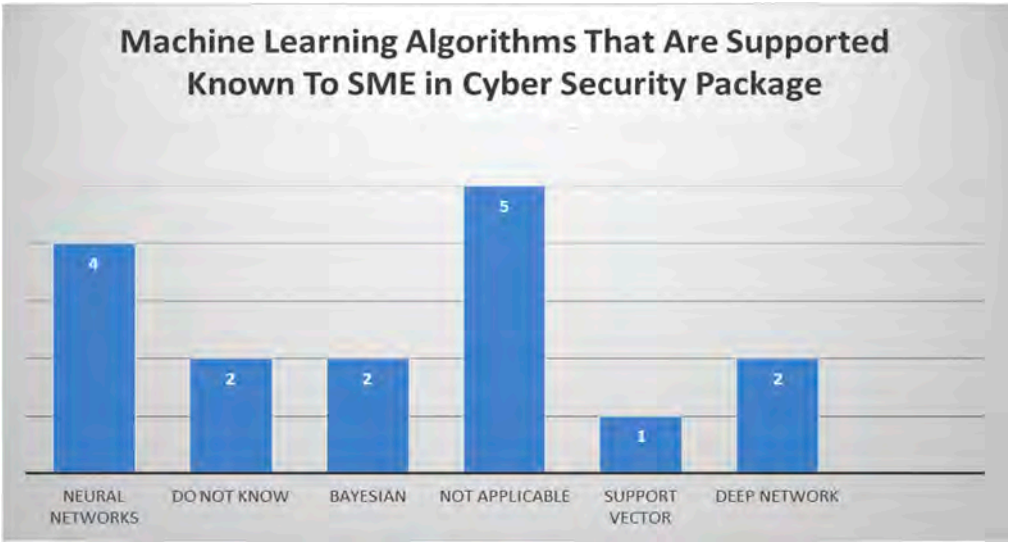


**Figure 10.** ML algorithms that are supported and known to SMEs in CS packages.

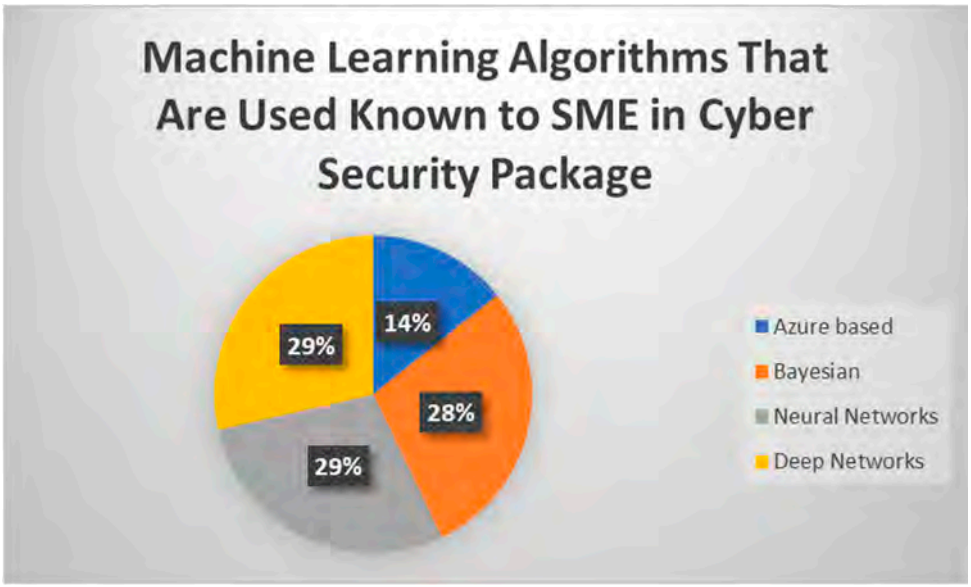


**Figure 11.** ML algorithms that are used and known to SMEs in CS packages.

Figure 12 shows this awareness and reveals that 75% of the participants surveyed did not know the price for their ML within their cyber security packages. Those participants that were aware made up 20% of the results, with 5% leaving an Unknown response to the price question. From the 20% pool of participants, the industries they were in were largely from the engineering, IT, and consultancy sectors, with additional logistics and supply chain, plus printing. This pool was a mixture of male and females coming from university degree qualifications. These also happened to be from categories listed as management and technical expertise. The 75% pool were represented by yet again a mixture of age and gender; however, the position and the education varied, utilizing the other categories listed for education and positions. "Unknown" only represented 5% of the participant pool that was non-technical, although having a position in management.

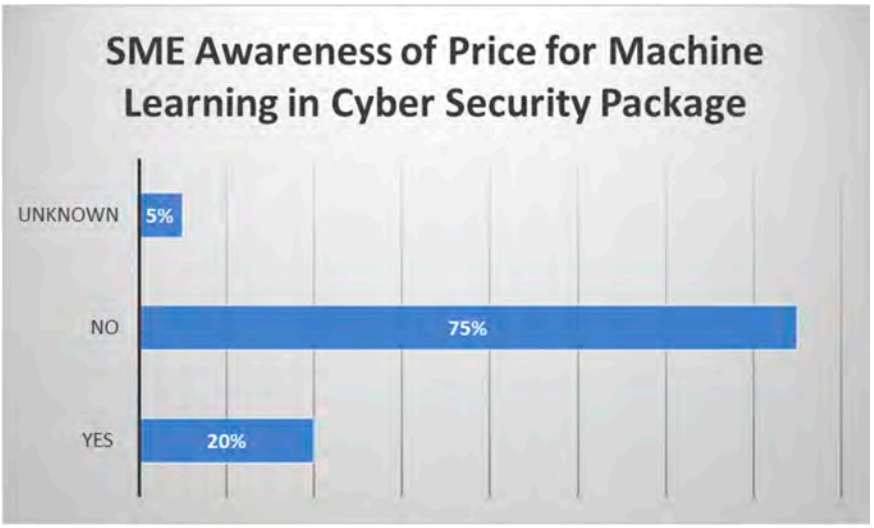


**Figure 12.** SME awareness of price for machine learning in cyber security packages.

## 5. Outcomes and Conclusions

Based on the analysis and findings, the study revealed that UK SMEs have the appropriate cyber security packages in place but are not necessarily aware of their full potential. Here there was a cross between price of these packages and what the full potential could cost, plus an understanding of the technical barriers within these selected UK SMEs. It also showed that management and their technical knowledge was not perhaps in depth to the level of ML and its algorithms. They were familiar with the security and safety of these packages given to their company; however, these SMEs could not identify further the technical aspects of these software. The SMEs were merely recommended these solutions from suppliers of these products through supply chain expertise and knowledge. Management wanted to learn more and have a better understanding of AI and ML, especially with the rise in cyber hacking due to the COVID-19 pandemic that resulted in staff and workers having to work from home, leaving data vulnerable to variables within the home networks and security. Some UK SMEs outsourced their IT and thus were not able to give a direct and true answer in this survey, hence relying on their internal technical expertise especially during the pandemic. The study also showed that management relied on their IT teams' expertise, but as determined from this survey, wanted to learn more about cyber security to personally understand how to protect their businesses. In many developed countries, most decision are made between SME management and their IT teams, and within this survey, it was shown that IT seems to be the main advisory team to understand the current benefits of new software and how these software can grow with the business and protect their data. The expertise of the IT team has proven important in how ML will be used in the future. Some participants did not know that AI and ML are built into the solution and expressed interest, while some showed disregard. Some did not think it was a high priority and felt that ML and cyber threats were not applicable to their business and not relevant to their industry. This is a useful point for policy to take note of, as awareness in the regulation of cyber security is important, especially in the context of SMEs and GDPR. The other point of importance was the costings of these software. Some UK SMEs judged cost to be too great for their operations.

It is apparent that awareness from SMEs in this pool of participants was poor, and that more work in raising awareness or being informed needs to take place. It could be that small ML e-learning packages in the form of video or slides targeted at UK SMEs might improve this awareness, emphasizing the importance of the subject to both the SMEs and policy, extending this to other developed nations. Further research should explore the development of these package with SMEs in mind. Here the study would like to see engagement between governing bodies such as GCHQ and NCSC coupled with GDPR.

This is very important, as both these governing bodies currently have a critical role in making the UK safe in cyber space. The important policy in the overall framework of GDPR suggests that emerging policy in the context of SMEs through education, training, and awareness should be emphasized.

The results highlight that although ML is a very effective technique for cyber security, adoption is poor amongst UK SMEs in those sampled in this paper, mainly due to cost and technical expertise. This paper highlights an important gap that can be fulfilled by perhaps more open-source and voluntary participants from the community to keep the UK SMEs safe. This article also highlights an important funding gap that could be fulfilled by the government to support SMEs in the form of grants, subsidies, and similar financial assistance through various public sector policies.

Whilst technology giants such as Google, Amazon, and Facebook might lead the path in its implementation of ML and cyber security, it is these high technology firms that will set precedence and bring awareness at the SME level and stress the importance of ML in keeping our cyber world safe. This has certainly been heightened even more from the cause and effect of the global COVID-19 pandemic, giving rise to the growing concern that is the cyber pandemic experienced in our time.

**Author Contributions:** Conceptualization, N.R. and A.J.; methodology, N.R. and A.J.; software, N.R. and A.J.; validation, N.R. and A.J.; investigation, N.R. and A.J.; writing—original draft preparation, N.R. and A.J.; writing—review and editing, N.R., A.J. and E.P.; visualization, N.R. and A.J.; supervision, A.J. and E.P.; funding acquisition, A.J. and E.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper has been supported by the KESS2, Knowledge Economy Skills Scholarships, Cardiff School of Technologies—Cardiff Metropolitan University and Aytel Systems Ltd., Cardiff, UK.

**Institutional Review Board Statement:** The study protocol was approved by the Ethics Committee of the Cardiff School of Technologies, Cardiff Metropolitan University.

**Informed Consent Statement:** All participants were informed about the study's objectives and gave their informed consent.

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| SME | Small and medium enterprises |
| MLCS | Machine learning cybersecurity |
| GDPR | UK General Data Protection Regulation |
| NIDS | Network Based Intrusion Detection System |
| AI | Artificial intelligence |
| ML | Machine learning |
| Brexit | Britain Exit |
| IoT | Internet of Things |
| IMD | International Institute for Management Development |
| SL | Supervised learning |
| USL | Unsupervised learning |
| RL | Reinforcement learning |
| DPA | Data Protection Act 2018 |
| SVM | Support vector machines |
| Bayesian | Naïve Bayes |
| PCA | Principal component analysis |
| C | Classification |
| DR | Dimensionality reduction |
| R | Regression |
| OE | Own experience |

| | |
|---|---|
| SM | Static models |
| OL | Own learning |
| CLR | Clustering |
| AWS | Amazon Web Services |
| CISO | Chief Information Security Officer |
| DoS | Denial-of-service |
| EIS | Enterprise immune system |
| FSB | National Federation of Self Employed & Small Businesses |
| ONS | Office of National Statistics |
| GHCQ | Government Communications Headquarters |
| NCSC | National Cyber Security Centre |
| SIEMs | Security Information and Event Management Systems |
| IDPS | Intrusion Detection and Prevention Systems |
| HIDS | Host-based |
| DDoS | Denial-of-service and distributed |
| SQLi | Structured query language |
| DGA | Domain generation algorithm |

## Appendix A. Literature Review Search Protocol

The research of cyber security within the UK SME market is growing in interest, and the approaches and applications to machine learning are also growing. This paper mainly consists of a literature review and research on some aspects of the barriers faced by UK SMEs for cyber security and machine learning approaches applied within the last five years. Relevant keyword searches were used from various search engines such as Google Scholar, Mendeley Data, ACM, and Scopus. In this paper, our main search engine used was Google Scholar, and keywords included "Barriers of Machine Learning in Cyber Security for UK SME" in various combinations, and search years were "All years" and narrowed down to the choice of year of "2017". Various combinations were attempted to narrow down the fields and to filter articles that were fit for purpose to this paper. Figures A1 and A2 both show various combinations of searches and advanced searches indicated in this paper.


**Figure A1.** Google Scholar search.

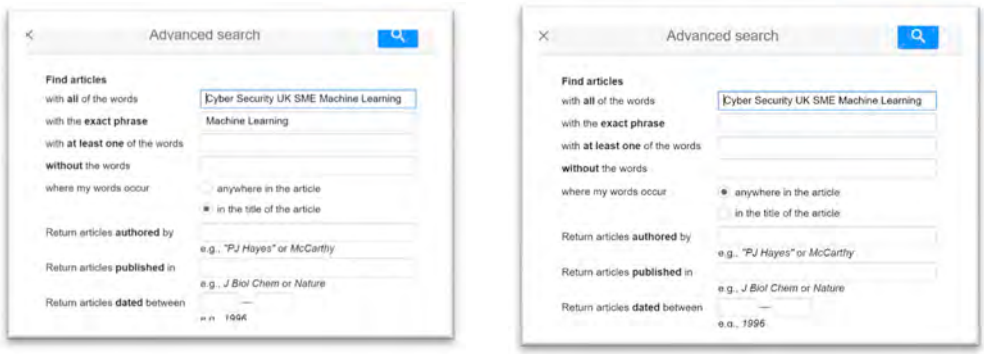


**Figure A2.** Advanced search with words fully in the title or anywhere in the article.

In accordance with these keywords above, the search came back with the following across various engines as shown in Table A1 below:

**Table A1.** Search engine keyword results.

| Search Engine | Keywords | Year | Number of Results |
|---|---|---|---|
| Google Scholar | Machine Learning Cyber Security UK SME | Anytime | About 19,700 results (0.08 s) |
| Google Scholar | Same as above | 5 years | About 16,800 results (0.13 s) |
| ACM | Same as above | 5 years | 141,303—anywhere |
| Met Search | Same as above | 5 years | 568 results |
| Mendeley Data | Same as above | Anytime | 191 results |

Depending on the keyword and the filtration, the search gave the results shown below in Figure A3. Looking for all the keywords in the title gave the results filtered to two documents.


**Figure A3.** Keyword search results after filtration.

Since the main search engine was Google Scholar, the same search was used, and the years shortened to the most current year of research.

**Table A2.** Search engine keyword results.

| Search Engine | Keywords | Year | Number of Results |
|---|---|---|---|
| Google Scholar | Barriers of Machine Learning in Cyber Security UK SME | Anytime | About 20,100 results |
| Google Scholar | Same as above | 2017 | About 17,300 results |
| Google Scholar | Same as above | 2018 | About 7140 results |
| Google Scholar | Same as above | 2019 | About 7600 results |
| Google Scholar | Same as above | 2020 | About 12,400 results |
| Google Scholar | Same as above | 2021 | About 4730 results |

The results were not succinct, as the word "barriers" was not in contest with the preferred keyword search and was embedded in the returned results. Removing the word produced the same results with no changes in returns.

By changing the keyword search to reflect a more focused return, the below syntax was used:

((“cyber security” SME network [Title/Abstract]) OR (cyber security SME network barriers [Title/Abstract])

The searches shown in Table A3 below used the syntax above to perform the search based on title and abstract for level of importance.

**Table A3.** Search engine keyword results.

| Search Engine | Keywords | Year | Number of Results |
|---|---|---|---|
| Google Scholar | ((“cyber security” SME network [Title/Abstract]) OR (cyber security SME network barriers [Title/Abstract]) | Anytime | About 69 results |
| Google Scholar | Same as above | 2017 | About 57 results |
| Google Scholar | Same as above | 2018 | 5 results |
| Google Scholar | Same as above | 2019 | 5 results |
| Google Scholar | Same as above | 2020 | About 47 results |
| Google Scholar | Same as above | 2021 | About 22 results |

Table A3 verified that there is a growing concern in the research area of SMEs choices for their cyber security and how machine learning could benefit or hinder and why barriers seem to be imposed when making these choices to move technology forward to keeping the data continually protected and secure.

## Appendix B. Qualtrics Questions

Q1—Please state your Industry

Q2—What is the highest level of education you have received?

Q3—What is your age range?

Q4—How do you identify?

Q5—Please state your Position in SME Company (tick all that applies)

Q6—Do you have Cyber Security Software Packages that Protects your business from Cyber Threats?

Q7—If NO or Unaware, please state reason.

Q8—If YES, Which Cyber Security Software Package do you use? Please tick all that is appropriate to your business.

Q9—Does your Cyber Security Software Package Support Machine Learning to Detect Cyber Attacks?

Q10—If NO or Unknown, please choose one of the below on the reasons for not using Machine Learning feature.

Q11—If YES, which Machine Learning Algorithm does the Cyber Security Software Package Support?

Q12—If YES, which Machine Learning Algorithm does the Cyber Security Software Package Use from the Options above If Known?

Q13—If YES, when did you start using the Machine Learning within the Cyber Security Software Package? Please specify date (month and year).

Q14—Please rate your satisfaction with the following in regard to your recent interaction with your IT team for the knowledge they bring for Cyber Security Packages and Machine Learning Algorithms.

Q15—Do you know the price of the Cyber Security Software Package WITHOUT its Machine Learning feature (Annually Per User)?

Q16—Do you know the price of the Cyber Security Software Package WITH its Machine Learning feature (Annually Per User)?

Q17—Please rate your satisfaction with the following in regard to your recent interaction with us if Pricing was known to you compared to the type of Cyber Security Software Package you had before?

Q18—How satisfied are you with the current Cyber Security Software Package you have in place?

Q19—Is there anything else you’d like to share that could make awareness of Machine Learning and its Algorithms better?

Q20—Would it be okay for us to follow up with you about your responses?
Q21—What's the best email address to reach you?

## References


1. Saleem, M. Brexit Impact on Cyber Security of United Kingdom. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
2. Industry 4 and the Pandemic. Available online: https://www.imd.org/research-knowledge/articles/the-pandemic-might-have-provided-a-clearing-in-the-woods-for-industry-4/ (accessed on 3 March 2021).
3. Warwick, K. *Artificial Intelligence: The Basics*; Routledge: Abingdon, UK, 2013.
4. Information Commissioner's Office (ICO)—SME Web Hub—Advice for All Small Organizations. Available online: https://ico.org.uk/for-organisations/sme-web-hub/ (accessed on 8 March 2021).
5. Robocop Machine Learning Expense Fraud. Available online: https://www.cbronline.com/emerging-technology/robo-cop-machine-learning-expense-fraud/ (accessed on 3 March 2021).
6. Technology New. Available online: https://www.bbc.co.uk/news/technology-30290540 (accessed on 3 March 2021).
7. Aslam, F.; Aimin, W.; Li, M.; Ur Rehman, K. Innovation in the era of IoT and industry 5.0: Absolute innovation management (AIM) framework. *Information* **2020**, *11*, 124. [CrossRef]
8. Hewage, C.; Jayal, A.; Jenkins, G.; Brown, R.J. A Learned Polyalphabetic Decryption Cipher. *SNE* **2018**, *28*, 141–148. [CrossRef]
9. IWM How Alan Turing Cracked the Enigma. Available online: https://www.iwm.org.uk/history/how-alan-turing-cracked-the-enigma-code (accessed on 3 March 2021).
10. Ghoseiri, K.; Nadjari, B. An ant colony optimization algorithm for the bi-objective shortest path problem. *Appl. Soft Comput.* **2010**, *10*, 1237–1246. [CrossRef]
11. What is Machine Learning. Available online: https://www.toolbox.com/tech/artificial-intelligence/tech-101/what-is-machine-learning-definition-types-applications-and-examples/#_003 (accessed on 3 March 2021).
12. Buczak, A.L.; Guven, E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun. Surv. Tutor.* **2015**, *18*, 1153–1176. [CrossRef]
13. Fraley, J.B.; Cannady, J. The promise of machine learning in cybersecurity. In Proceedings of the SoutheastCon, Concord, NC, USA, 30 March–2 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [CrossRef]
14. Arora, A.; Jayal, A.; Gupta, M.; Mittal, P.; Satapathy, S.C. Brain Tumor Segmentation of MRI Images Using Processed Image Driven U-Net Architecture. *Computers* **2021**, *10*, 139. [CrossRef]
15. Lee, K.; Caverlee, J.; Webb, S. Uncovering Social Spammers: Social Honeypots + Machine Learning. In Proceedings of the SIGIR'10, Geneva, Switzerland, 19–23 July 2010.
16. Uppal, S.; Jayal, A.; Arora, A. Pairwise Reviews Ranking and Classification for Medicine E-Commerce Application. In Proceedings of the 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 8–10 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
17. LAZIĆ, L. Benefit from Ai in cybersecurity. In Proceedings of the 11th International Conference on Business Information Security (BISEC 2019), Belgrade, Serbia, 18 October 2019.
18. Five Top Machine Learning Use Cases for Security. Available online: https://www.csoonline.com/article/3240925/5-top-machine-learning-use-cases-for-security.html (accessed on 3 March 2021).
19. How Credit Card Companies are Fighting Cyber Frauds. Available online: https://cio.economictimes.indiatimes.com/news/digital-security/heres-how-visa-mastercard-and-paypal-are-fighting-cyber-frauds-with-ai/79381050 (accessed on 3 March 2021).
20. Vähäkainu, P.; Lehto, M. Artificial intelligence in the cyber security environment. In Proceedings of the ICCWS 2019 14th International Conference on Cyber WarfarSe and Security: ICCWS, Stellenbosch, South Africa, 28 February–1 March 2019; p. 431.
21. Amazon Web Services, Inc. Amazon Macie FAQ. *Amazon.* 2018. Available online: https://aws.amazon.com/macie/faq (accessed on 3 March 2021).
22. Proko, E.; Hyso, A.; Gjylapi, D. Machine Learning Algorithms in Cyber Security. In *RTA-CSIT*; 2018; pp. 203–207. Available online: https://www.semanticscholar.org/paper/Machine-Learning-Algorithms-in-Cyber-Security-Proko-Hyso/67525df429c50af9ae5fe10949cd7d279ee1184f (accessed on 27 October 2021).
23. Orche, A.E.; Bahaj, M. Approach to Combine an Ontology-Based on Payment System with Neural Network for Transaction Fraud Detection. Available online: https://astesj.com/v05/i02/p69/ (accessed on 27 October 2021).
24. Tech Giants Using AI against Hackers. Available online: https://analyticsindiamag.com/how-tech-giants-like-amazon-microsoft-google-are-using-ai-against-hackers/ (accessed on 3 March 2021).
25. Hackers Trick Tesla. Available online: https://www.technologyreview.com/2019/04/01/65915/hackers-trick-teslas-autopilot-into-veering-towards-oncoming-traffic/ (accessed on 3 March 2021).
26. Ford, V.; Siraj, A. Applications of machine learning in cyber security. In Proceedings of the 27th International Conference on Computer Applications in Industry and Engineering, New Orleans, LA, USA, 13–15 October 2014; IEEE Xplore: Kota Kinabalu, Malaysia, 2014; Volume 118.
27. Jayal, A.; McRobert, A.; Oatley, G.; O'Donoghue, P. *Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performance*; Routledge: Abingdon, UK, 2018.

28. UK Small Business Statistics, F.S.B. The Federation of Small Businesses. Available online: https://www.fsb.org.uk/uk-small-business-statistics.html (accessed on 1 September 2021).
29. SME Action Plan. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961722/SME-Action-Plan.pdf (accessed on 1 September 2021).
30. The Impact of the Coronavirus so Far: The Industries that Struggled or Recovered—Office for National Statistics (Ons.Gov.UK). Available online: https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/theimpactofthecoronavirussofartheindustriesthatstruggledorrecovered/2020-12-09 (accessed on 1 September 2021).
31. O'Leary, D.E. 'Big Data', the 'Internet of Things', and the 'Internet of Signs'. *Intell. Syst. Account. Financ. Manag.* **2013**, *20*, 53–65. [CrossRef]
32. Cox, M.; Ellsworth, D. Managing Big Data for Scientific Visualization. ACM Siggraph: USA. 1997. Available online: https://www.researchgate.net/publication/238704525_Managing_big_data_for_scientific_visualization (accessed on 27 October 2021).
33. Murtarelli, G.; Gregory, A.; Romenti, S.A. Conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *J. Bus. Res.* **2020**, *129*, 927–935. [CrossRef]
34. GHCQ Overview. Available online: https://www.gchq.gov.uk/section/mission/overview (accessed on 3 March 2021).
35. Intelligent Security Tools. Available online: https://www.ncsc.gov.uk/collection/intelligent-security-tools (accessed on 3 March 2021).
36. Small Business Guide: Cyber Security. Available online: www.ncsc.gov.uk (accessed on 3 March 2021).
37. López, M.Á.; Lombardo, J.M.; López, M.; Alba, C.M.; Velasco, S.; Braojos, M.A.; Fuentes-García, M. Intelligent Detection and Recovery from Cyberattacks for Small and Medium-Sized Enterprises. *Int. J. Interact. Multimed. Artif. Intell.* **2020**, *6*, 55–62. [CrossRef]
38. Rawindaran, N.; Jayal, A.; Prakash, E.; Hewage, C. Cost Benefits of Using Machine Learning Features in NIDS for Cyber Security in UK Small Medium Enterprises (SME). *Future Internet* **2021**, *13*, 186. [CrossRef]
39. Elmrabit, N.; Zhou, F.; Li, F.; Zhou, H. Evaluation of machine learning algorithms for anomaly detection. In Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, 15–19 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
40. Gupta, A.; Gupta, R.; Kukreja, G. Cyber Security Using Machine Learning: Techniques and Business Applications. In *Applications of Artificial Intelligence in Business, Education and Healthcare*; Springer: Cham, Switzerland, 2021; pp. 385–406.
41. Van Haastrecht, M.; Yigit Ozkan, B.; Brinkhuis, M.; Spruit, M. Respite for SMEs: A Systematic Review of Socio-Technical Cybersecurity Metrics. *Appl. Sci.* **2021**, *11*, 6909. [CrossRef]
42. Hiatt, J. *ADKAR: A Model for Change in Business, Government, and Our Community*; Prosci: Fort Collins, CO, USA, 2006.
43. DIGITAL SME Alliance. *The EU Cybersecurity Act and the Role of Standards for SMEs—Position Paper*; Technical Report; European DIGITAL SME Alliance: Brussels, Belgium, 2020.
44. Tam, T.; Rao, A.; Hall, J. The Good, The Bad and The Missing: A Narrative Review of Cyber-security Implications for Australian Small Businesses. *Comput. Secur.* **2021**, *109*, 102385. [CrossRef]
45. Qualtrics [Online Software]: Provo, UT, USA. Available online: www.qualtrics.com (accessed on 3 March 2020).
46. Acharya, A.S.; Prakash, A.; Saxena, P.; Nigam, A. Sampling: Why and how of it. *Indian J. Med. Spec.* **2013**, *4*, 330–333. [CrossRef]
47. Greenfield, T. *Research Methods for Postgraduates*; Arnold: London, UK, 2002.
48. Sarantakos, S. *Social Research*; Macmillan: Basingstoke, UK, 2013.
49. Kirby, M.; Konbel, F.; Barter, J.; Hope, T.; Kirton, D.; Madry, N.; Manning, P.; Trigges, K. *Sociology in Perspective*; Heinnemann: Oxford, UK, 2000.