HITTITE JOURNAL OF SCIENCE AND ENGINEERING

e-ISSN: 2148-4171
Volume: 12 • Number: 1
March 2025

Determining the Cyber Risk Matrix and Actions Created by Company Employees with Machine Learning

Esma Sığırtmaç[1] | **Musa Balta**[1] | **Deniz Balta**[2]
[1]Sakarya University, Department of Computer Engineering, Sakarya, Türkiye.
[2]Sakarya University, Department of Software Engineering, Sakarya, Türkiye.

Corresponding Author
Esma Sığırtmaç
E-mail: esma.sigirtmac@ogr.sakarya.edu.tr Phone: +90 543 227 35 87
RORID: https://ror.org/04ttnw109

Article Information
Article Type: Research Article
Doi: https://doi.org/10.17350/HJSE19030000346
Received: 30.01.2024
Accepted: 09.12.2024
Published:25.03.2025

Cite As
Sığırtmaç E, et al. Determining the Cyber Risk Matrix and Actions Created by Company Employees with Machine Learning. Hittite J Sci Eng. 2025; 12(1):1-14.

Peer Review: Evaluated by independent reviewers working in at least two different institutions appointed by the field editor.
Ethical Statement: Not available.
Plagiarism Checks: Yes - iThenticate
Conflict of Interest: Authors declare no conflict of interest.

CRediT AUTHOR STATEMENT
Esma Sığırtmaç: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Supervision, Writing – review and editing. **Musa Balta:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft & editing. **Deniz Balta:** Conceptualization, Data Curation, Formal Analysis, Methodology, Validation, Writing.

Determining the Cyber Risk Matrix and Actions Created by Company Employees with Machine Learning

Esma Sığırtmaç[1,*] | Musa Balta[1] | Deniz Balta[2]

[1]Sakarya University, Department of Computer Engineering, Sakarya, Türkiye.
[2]Sakarya University, Department of Software Engineering, Sakarya, Türkiye.

Abstract

In today's digital age, the integration of various fields with the internet and technology has enabled people to meet many issues online, from their basic needs to business, banking and entertainment. However, this digital transformation poses new threats for companies, especially in terms of cyber security. Cyber-attacks can directly harm companies, disrupting systems and damaging their credibility. Despite taking technical measures, companies often encounter weaknesses due to the human factor. This study aims to identify profiles that may cause security vulnerabilities and increase the company's cybersecurity defense level with appropriate actions. When the results are examined, it is discovered that people with a certain experience range have the same approaches. Using K-means and Mean Shift clustering algorithms, individuals are grouped according to their behaviors and a cyber risk matrix is created for the company, and it is determined which situations these people fall into which risk category. As a result of the data obtained, it is clearly seen that the human factor has emerged as a more important issue than the technical dimension in cyber security.

Keywords: Security, Machine Learning, Cyber Risk Matrix, Human Factor, Cyber Security Awareness

INTRODUCTION

Humanity, especially after Covid-19, has increasingly turned to using the internet for many tasks. New applications and systems have been developed, and many processes have been moved to the virtual environment. According to the report published by ENISA (The European Union Agency for Cybersecurity) on October 20, 2020, an increase in cyber-attacks and their varieties has been observed [1]. The closures and economic fluctuations experienced after Covid-19 have provided an opportunity for financially motivated criminals to target many corporate and institutional areas by exploiting the increased use of the internet. Tonya Ugoretz, Deputy Assistant Director of the Cyber Division of the FBI (Federal Bureau of Investigation), stated in 2020 that they used to receive 1000 cyber-attack complaints every day before Covid-19. However, after Covid-19, this number experienced a surge, reaching between 3000 and 4000 per day [2].

Research has observed an increase in cyber-attacks and their diversity. Conclusions have been drawn that companies need to take additional security measures to protect against cyber-attacks. However, according to statements from ENISA, existing measures may not be sufficient as attack methods evolve [1].

No matter how many mitigation techniques are taken, the error rate increases when humans are involved. According to reports from ENISA, the most commonly used attack methods during the Covid-19 period are phishing, social engineering, malware, misconfiguration, poor policies, and technology-induced security vulnerabilities [1]. This ranking is made from the most used to the least used.

Organizations should place significant emphasis on technological solutions to combat potential cyber threats. Recent research in cybersecurity strongly agrees that a holistic approach is necessary to resist cyber-attacks, in contrast to relying solely on technical solutions. This is particularly noticeable in well-targeted sectors like education and health, as well as emerging fields such as autonomous vehicles. User behaviors and attitudes can undermine technological advancements.

Due to these developments and the increasing use of the internet, the presence of people of all ages online has become a significant danger for companies [3-4]. Despite companies investing in and prioritizing technical measures, undesirable situations can arise due to the carelessness of an employee within the company.

There are studies in the literature to increase cybersecurity awareness and consciousness. In their study, Avcı and Oruç (2022) examined the relationship between university students' information security awareness and cybersecurity behaviors according to various demographic variables. In order to increase students' awareness, solution suggestions such as including relevant courses in the curriculum, informing students about these issues from an early age, and making them aware of the importance of ensuring cybersecurity were presented [5]. In the study of Yiğit and Seferoğlu (2019), university students' cybersecurity behaviors were examined according to personality traits and variables such as gender, grade level, department, information security training status, and weekly internet usage time. At the end of the study, in the light of the findings, it was suggested that cybersecurity training should be emphasized and students' personality traits should be taken into account in these trainings [6]. In their study, Yetgin and Karakaya (2020) measured the personal cybersecurity perceptions of academic and administrative staff working at Karabük University. The data collected with the survey method were analyzed with Cronbach Alpha, single sample t test, independent sample t test, and ANOVA test. It is not stated that there are differences in the perceptions of employees about personal cyber security according to the parameters and various training suggestions are given [7]. Gündüz and Das (2022) mentioned in their study that personal cyber security awareness can be increased on the end user side with cyber awareness. The article suggests new approaches for end users to create secure passwords within the scope of ensuring the security of online individual identity data [8]. Tokmak (2023) determined the cyber security awareness levels of students about cyber threats with machine learning methods in his study. Data was collected with the survey method. The effect of factors such as the department the students study and gender on the cyber security awareness of students was emphasized [9]. In the

study conducted by Cam et al. (2019), the Internet usage levels and personal information security attitudes of students, employees and academicians at Gümüşhane University were examined. Exploratory factor analysis, descriptive statistical analyses and two-way variance analysis were used for the analysis of the data. The research results emphasized that studies should be conducted to increase the information security awareness levels in higher education institutions. [10].

Based on the data obtained, it is clearly seen that the human factor has become more crucial in cybersecurity than the technical dimension. Within the scope of this study, the aim is to minimize the human risk factor in the field of cybersecurity to the lowest possible level. Investigations have been conducted in various areas such as business life, personal life, education, etc., where the internet can be utilized. In order to minimize the human factor, specific methods and research results have been consolidated at a common point, leading to a human-centric approach.

Due to evolving new cyber-attack methods and threats, it is essential to raise awareness among individuals. Tailored education needs to be provided to individuals based on their specific needs, guiding them through instructive actions and actions to avoid. Research and observations have revealed that technical security measures alone are not sufficient.
In this study, individuals are aimed to be grouped based on survey results. Tests have been conducted on various parameters and data using technologies, leading to similar results.

The literature on cyber risks in companies has indeed explored various methodologies for grouping individuals based on survey results, particularly in relation to risk assessment and management. Several studies have examined the factors contributing to cyber risks, the effectiveness of cyber insurance, and the implications of organizational behavior in mitigating these risks. One significant study by Talesh discusses how cyber risk management services, including cyber insurance, not only reduce risks but also shape compliance behaviors within organizations. This research highlights the role of insurance companies as "compliance managers," indicating that organizations are increasingly institutionalizing responses to cyber risks through insurance policies [11]. This perspective is critical as it suggests that organizations can be grouped based on their compliance strategies and the extent to which they engage with cyber insurance. This aligns with findings from Kenny et al., who identified specific demographic factors that correlate with cyber-victimization among different groups [12]. Such demographic insights could be crucial for grouping individuals based on their risk profiles. In the context of cybersecurity awareness, Tempestini et al. developed a tool to assess cybersecurity knowledge among college students, categorizing participants into risk groups based on their reported behaviors and experiences [13]. This method of grouping individuals based on survey responses is particularly relevant to the task of identifying cyber risk profiles.

Moreover, the work of Bergh and Junger reviews victim surveys related to cybercrime across Europe, emphasizing the need for standardized methodologies in assessing cyber risks. They argue that such standardization can facilitate better grouping of organizations based on their experiences with cybercrime victimization [14]. This aligns with the idea that organizations can be categorized based on their risk profiles and responses to cyber incidents. In the context of cyber incident prediction, Pramoda et al. present a novel model that utilizes machine learning to assess the risk of cyber incidents among different demographics. Their findings indicate that increased internet usage correlates with a higher likelihood of cyber incidents, suggesting that organizations can be grouped based on their employees' internet usage patterns and associated risks [15]. This quantitative approach to risk assessment is crucial for developing targeted interventions. Additionally, the research by Nurse et al. emphasizes the importance of understanding the complexities of assessing security risks in Internet of Things (IoT) systems. Their findings indicate that professionals from various sectors identify key issues in cyber-risk assessment, which can inform how organizations are grouped based on their technological vulnerabilities and risk management practices [16]. This highlights the necessity of a multidisciplinary approach to cyber risk assessment, which can lead to more effective grouping of organizations based on their specific risk profiles. Furthermore, the study by Cains et al. focuses on defining cyber security and cyber security risk within a multidisciplinary context, utilizing expert elicitation methods. This research underscores the importance of a common understanding of cyber risks, which can facilitate the grouping of organizations based on their perceived vulnerabilities and risk management strategies [17].
In summary, the literature provides substantial evidence that grouping individuals and organizations based on survey results related to cyber risks has been explored through various lenses, including compliance behaviors, victimization surveys, incident prediction models, and multidisciplinary definitions of cyber security. These studies collectively contribute to a deeper understanding of how organizations can be categorized based on their cyber risk profiles and management strategies.

Upon reviewing the results, it was discovered that individuals within certain experience year ranges exhibit similar approaches. Accordingly, individuals within certain experience year ranges show similarities in terms of mistakes and shortcomings. By creating a cyber risk matrix, it has been determined which risk category corresponds to these individuals for specific situations, and necessary precautions and actions have been recommended.

By using the cyber risk matrix, actions are taken for existing employees or new incoming employees based on their position on the matrix. Through the cyber risk matrix, the deficiencies in employees' cybersecurity aspects are addressed, aiming to minimize the human factor in cyber threats.

MATERIAL AND METHODS

Determining the Algorithm

There are many questions and answers related to the study. However, how individuals will behave is uncertain. The use of

machine learning is important for clustering individuals, but when it comes to human behavior, a clear result cannot be obtained. Therefore, using any form of supervised machine learning algorithm is not considered.

In the early stages of the study, clustering with regression algorithms was attempted. The analysis of behavior was based on the measures individuals took and whether they had previously experienced a cyber-attack. Behavior analysis was conducted based on this information. However, in the survey results, it was observed that a significant number of individuals claimed to have never experienced a cyber-attack or did not know about it. Therefore, a clear conclusion about whether individuals have experienced a cyber-attack or not could not be reached. Additionally, a person who has previously experienced a cyber-attack is likely to have learned from the incident and is less likely to be targeted again.

As a result of observations and investigations, it was concluded that supervised studies would not yield satisfactory results. Due to the inherent lack of clear results in human behavior, using an unsupervised algorithm would be more appropriate when working with these individuals. Upon examining unsupervised algorithms, it was found that there are various types available. The investigations revealed that clustering algorithms in the category of unsupervised algorithms provided the desired results.

Clustering involves algorithms that group data based on similarities according to entered parameters. These algorithms have different working principles, such as distance to the center, distance to neighbors, etc. However, they all share a common point: grouping similar data.

K-Means Algorithm

To test the functionality of the code and the established system, testing was initially started with K-means. Because the K-Means algorithm is a frequently used method in cluster analysis of data expressed with high-dimensional and continuous variables, especially survey data. In the evaluation of cyber awareness survey results, the K-Means algorithm is an effective method to separate individuals into similar groups according to their awareness levels [18].

The K-Means algorithm is a center-based clustering method used to divide data into K clusters. The algorithm works on the assumption that each cluster is clustered around a center point (centroid) and that data points with similar characteristics are placed in the same cluster. Each data point is assigned to the nearest center by measuring the distances to the center points, so that similar data points are included in the same cluster. The Working Steps of the K-Means Algorithm are as follows.

Determining the K Value: First, it is necessary to determine how many clusters will be created (K value). The K value is usually determined according to the structure of the data or the purpose of the analysis.

Assigning Initial Centers: K random center (centroid) points are selected.

Assignment Step: Each data point is assigned to the nearest center point.

Updating Centers: New center points are determined by calculating the average for each cluster.

Iteration: Assigning data points to clusters and updating the centers is repeated until the centers do not change or a specified number of iterations is reached.

In this study, the elbow method was used to find the appropriate K value in areas where K-means was used. When the groups were examined in order from one to nine, it was seen that the break in the resulting graph occurred at 4. In this case, the K value was determined as 4 in the algorithm. The number of iterations was determined as 300 by default.

Mean Shift Algorithm

The K-means algorithm works with two parameters. In some complex cases, K-means is sufficient. However, in more problematic and complex cases, sufficient results are not obtained. For this, it is necessary to use a new structure that takes three parameters with the same working method and the same grouping system. Data is analyzed by switching between K-means or Mean Shift according to the need [19].

Mean Shift is based on the principle of shifting cluster centers towards the areas where the densities are highest. This algorithm does not require any fixed K number (predetermined number of clusters), instead it creates clusters by itself by focusing on the regions where the data density is. Thanks to this feature, it is ideal for revealing natural clusters in the data structure. The Working Steps of the Mean Shift Algorithm are as follows.

Determining the Starting Points: Each data point is initially considered a cluster center (centroid).

Mean Shift: Each point is “shifted” toward the center of the surrounding data density. A mean vector is calculated for each point by considering the other points within a certain bandwidth.

Approaching the Density Peaks: All data points continue this shifting process iteratively and eventually cluster at the density peaks. This process continues until a density center is found where the centers do not change any further.

Creating the Clusters: Once the shifting process is complete, the centers that are close to each other are merged, thus obtaining clusters.

COLLECTION AND FORMATTING OF DATA

Survey Content and Questions

In this study, a survey called Cyber Awareness Form was created for company employees. Participants access the Cyber Awareness Form survey via the internet and fill out the survey anonymously. The Cyber Awareness Form survey consists of 22 questions. These questions aim to measure

the cyber awareness rate of company employees. The Cyber Awareness Form survey was mostly filled out by employees in sectors such as IT, banking, finance, automotive, etc. A total of 659 people responded to the survey called Cyber Awareness Form. The distribution of people by sector is shown in Figure 1.

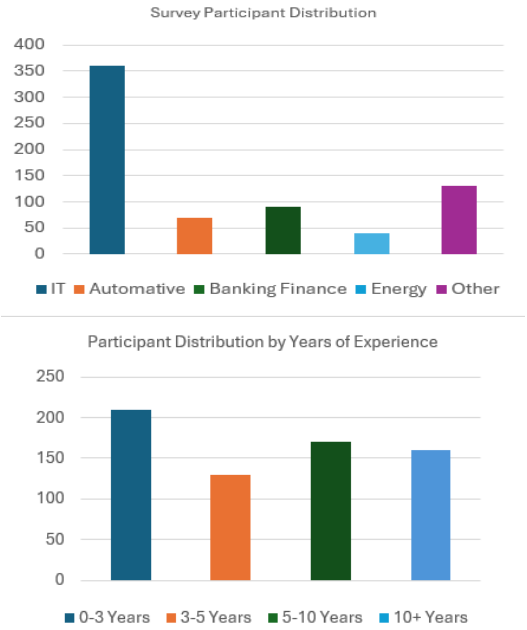

Figure 1. Participant distribution by sector and years of experience

Password security refers to the measures taken to increase the security of passwords used to access digital accounts. A password is an authentication information that is often used in conjunction with a username and is kept confidential to ensure account security. The main goal is to keep user accounts safe by ensuring that these passwords are protected against unauthorized access.

The survey also includes password security questions. Passwords that are long, complex and consist of random characters should be preferred. The questions aim to measure the user's password security knowledge by asking what kind of characters the passwords contain, the frequency of changing passwords, and how users store their passwords.

Email security refers to the measures taken to protect electronic communication from various threats. These measures aim to enhance the privacy, integrity, and security of messages sent and received through email services. Email security plays a critical role in safeguarding sensitive information for individuals, businesses, and organizations, as well as in resisting cyber-attacks and establishing reliable communication channels.

Email security involves implementing measures to protect electronic communication and mitigate threats such as unauthorized access, data leakage, phishing, and malicious software. Employees should carefully verify the links in the emails they receive. Reporting harmful or fraudulent emails to the relevant team is crucial for the company's security.

Testing and Performance Measurement

First of all, the textual data obtained from the surveys were converted into numerical data and made ready to be used in algorithms. After that, random data was created in certain models to determine the most accurate algorithm. Apart from the survey, a study needs to be conducted to see the performance and outputs of the most well-known algorithms with randomly generated data. A structure has been prepared in which 10 algorithms can be tested. With the generated random data, spiral, circular, ring, linear and random etc. 6 types of data types have been prepared: The prepared data were entered into the algorithms one by one and the graph in Figure 2 was obtained.

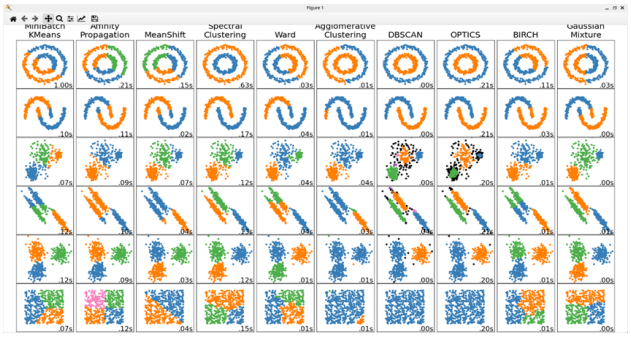

Figure 2. Algorithm comparison chart.

When examining the graph in Figure 2, several algorithms deemed suitable for use have been identified. Through research on these algorithms, the goal is to progress using these algorithms. The first of the identified algorithms is DBSCAN. The main advantages of this algorithm include not requiring the pre-specification of the number of groups, ease of clustering complex and varied data, and the presence of the concept of noise [4]. Due to these specified features of the DBSCAN algorithm, its usage has been observed to be appropriate. Especially, the concept of noise will be useful for exceptional cases outside the groups. However, over time, in some cases, all points from survey data have been perceived as noise. The use of the DBSCAN algorithm was deemed inappropriate due to considering all survey responses as noise.

After understanding the DBSCAN situation, other algorithms similar to it were ruled out. As a result of research, the decision was made to continue with the K-means algorithm for the data sets of our cybersecurity awareness survey. Finally, when exploring the Mean Shift algorithm, which produces similar outputs to K-means, it was noticed that it can take 3 parameters. In this case, it was identified that by adding another parameter, Y, instead of just adding the experience years and X as parameters, a three-dimensional graph can be plotted.

For example, by adding 3 parameters such as experience years, those who have fallen victim to phishing attacks, and mail URL (Uniform Resource Locator) check, we can make an inference about individuals' awareness levels. The use of the Mean Shift algorithm will be necessary to establish the structure due to its ability to take 3 parameters. Additionally,

it can yield good results in a complex dataset.

As seen in Figure 2, the data that needs to be examined will appear randomly distributed on a flat plane, such as the ones in the 3rd or 5th shape on the graph. Clustering distribution could not be done smoothly with K-means, Mean Shift, and a few other algorithms. When Figure 2 is examined, it is observed that while some data should be divided into at least three groups, some are divided into two groups and some into one group. Also, there are algorithms that perceive the dataset as noise. Assuming that the survey data is not so complex, the use of algorithms that do not separate into at least three groups will not be appropriate.

After the decision, instead of test data or random data; The formatted version of the survey data will be tested on these algorithms. After the data was run, all DBSCAN data was detected as noise. DBSCAN perceived all questions with two answers as noise. For questions with more than two answers, only a single data group was created. It is seen that using the DBSCAN algorithm within the scope of this study will not give accurate results.

After the test with K-means, appropriate results were obtained for questions with two answers. As seen in Figure 3 the center points have been removed for a question with a yes or no answer. In addition, the grouping process was carried out in line with the needs.

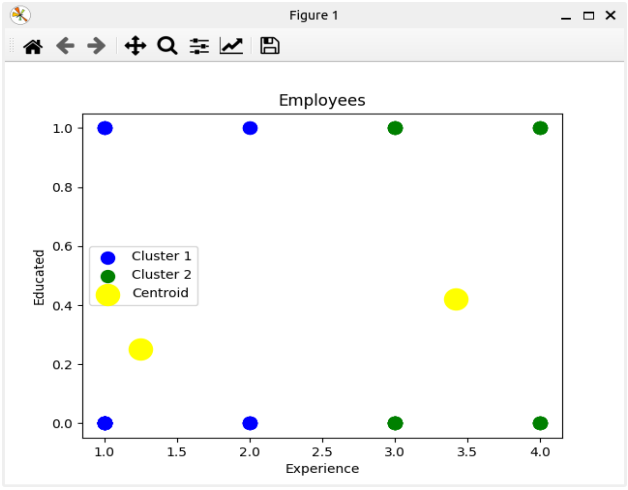

Figure 3. K-means plot tested with survey questions.

An example parameter has been selected for the accuracy of the algorithm, and the accuracy of the graph generated in Figure 3 has been tested with this parameter. According to the description in the graph, the dataset is divided into 2 groups. One group, represented by the blue color, includes employees with experience intervals of 0-3 and 3-5 years. The other group, shown in green, represents those with experience intervals of 5-10 and 10+ years. Another parameter is whether cybersecurity awareness training has been received or not. Looking at the centroids (centers of mass), the centroid for the blue group is located at the level of 0.2. The centroid for the green group is found at the level of 0.5. According to the obtained data, it is observed that individuals with fewer years of experience mostly have not undergone cybersecurity awareness training.

When the filled survey data is filtered and examined, it is likely that the situation appears this way. Upon examining the dataset, it is observed that the education level of individuals with 0-5 years of experience is lower than those with more than 5 years of experience. Additionally, in the group of participants with 0-5 years of experience, the number of individuals with 0-3 years of experience is observed to be higher than the group with 3-5 years of experience. The center of the blue group in the graph is also seen to shift to the left because the number of participants with 0-3 years of experience is higher as the center of mass. This test has been further validated with a few more data points, and similar studies on Mean Shift graphs have demonstrated effective clustering, indicating the accurate functioning of the algorithms.

RESULTS AND DISCUSSION

After ensuring the accuracy of the graphs, the next stage is the examination and interpretation of the data. In this stage, the progress was as follows: Other data were run according to the main parameter, which is the years of experience, and the situation was noted. During the processes, responses for different years of experience were compared with each other. After the general processes, the responses fed into the algorithm were classified based on the years of experience. For example, the algorithm was run for individuals with only 0-3 years of experience, and detailed data were examined. Then, a similar analysis was conducted for individuals with 3-5 years of experience.

As seen in Figure 3, there is a shift to the left at the center of the blue group. This means that there are more individuals with 0-3 years of experience in the blue group. Individuals within the 0-3 years of experience range determine the position of the center point and influence the responses of individuals with 3-5 years of experience. Individuals with 0-3 years of experience form the majority in the group representing the 0-5 years of experience range. This would lead to incorrect results, so after running the algorithm in a general sense, detailed analyses were conducted for each experience year group.

For the subsequent processes, the dataset was interpreted with different parameters, and a matrix was created. For this, a structure moving from general to specific was established. In this structure, individuals of all ages and experiences were examined under a single framework. At the end of the examination, it is more clearly understood whether individuals pose a cyber risk based on the study conducted by age. This results in a conclusion about under what conditions individuals create risks.

The first examined data is whether individuals have experienced a cyber-attack before. Since the answer to this question is more than two, the Mean Shift algorithm has been used to obtain the most accurate result. Three parameters, namely age, years of experience, and whether they have experienced a cyber-attack before, were provided to the

algorithm. When Figure 4 is examined, it can be seen that the data specifying years of experience is located at the bottom of the graph.

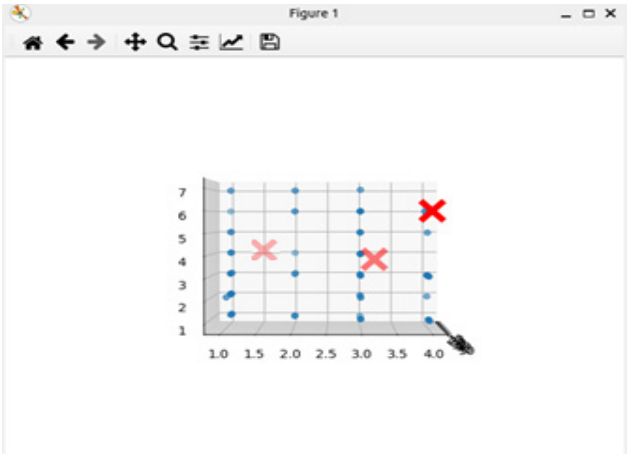

Figure 4. Previous cyber-attack incident graph.

In the graph in Figure 4, the data densely packed on the right side represents the age of individuals. The data on the left side represents the response individuals gave to this question. Since there are more than two answers to the question, the data needs to be formatted. According to the numbers on the left side of the graph, the answers are as follows: 1 = "I don't know," 2 = "I haven't experienced it," 3 = "Compromise of information due to link redirection via email (Phishing)," 4 = "Navigating on harmful websites," 5 = "Use of familiar disks/CDs, etc. used on different devices," 6 = "Through an application downloaded from the internet," 7 = "Through zip/rar files obtained from a third party."

In the interpretation phase of the graph in Figure 4, the centers with red crosses and blue dots constitute the majority of the answers. The points where the majority are present are examined and the data is recorded. When Figure 4 is examined, it can be seen that people with 0-3 and 3-5 years of experience are mostly grouped in the fourth answer. This means that people who have just started working life and have little experience generally choose the option number 4, "As a result of browsing harmful sites". People with experience between 0-3 years may not have checked the HTTPS (Hypertext Transfer Protocol Secure) on the sites they visit, or they may not have checked the reliability of any site they do not know, even if it has HTTPS. While browsing these sites, people may have clicked on an ad or downloaded an application, file, etc. They may have downloaded. For this reason, they may have caused malicious software to be installed on the computer, their cookies to be stolen, and their personal information to be stolen. In this case, it is concluded that people with little experience should be careful while surfing the Internet. For this reason, companies should raise awareness among their employees on this issue. People should be given training and seminars about safe internet browsing. Companies can prepare traps so that people can learn about the event by allowing them to experience the event. In the final stage of my study, these issues are mentioned among the actions that need to be taken.

When the data is examined in detail, it can be seen that people with less experience mostly choose the "I don't know" option. This is an indication that people with little experience act unconsciously. This process is also done based on years of experience and age. Detailed representations are also available for review.

A variation of the same situation is observed in people with more years of experience. It can be seen that similar cyber incidents occur as a result of applications downloaded from websites (number 6). Even though the site is safe, you should not download it. After downloading, an inactive virus may enter the computer and spread across the network, leaving the door open. Therefore, one should acquire the habit of being careful while surfing the Internet. For the next analysis, parameters were changed, and the information about whether individuals have received cybersecurity training was analyzed. As seen in Figure 5.a, when looked at in general, the number of those who have received training is observed to be low. Especially among individuals with less experience, the number of those who have received training is observed to be low. The graph and data in Figure 5.a have been examined. The question asked is whether they have received cybersecurity training before, and it consists of yes/no answers. When these answers are coded, resulting in 1 and 0, it is noticed that the green and blue groups only consist of 1s and 0s.

Since the dataset does not contain complex responses, the use of the Mean Shift algorithm is not appropriate. When the dataset is processed with the K-means algorithm, a correct result is obtained.

As seen in Figure 5.a, the algorithm has created 2 groups. The formed groups are divided into individuals with 0-5 years and more than 5 years of experience. After making general interpretations, a detailed examination can be conducted to obtain a more accurate result. Individuals with experience levels of 0-3 years and 3-5 years (numbers 1 and 2) show a high proximity to zero. These individuals pose a risk to the company. Additionally, when this result is combined with the information seen in Figure 4, it is observed that individuals with less than 5 years of experience often respond 'I don't know' to the question of whether they have experienced a hacking incident before. If this information is combined with not having received cybersecurity training, companies should consider providing cybersecurity training from the beginning based on the experience years of new or existing employees. If we examine Group 2, it is observed that this ratio is halved. Detailed examinations will be conducted in the later stages of the study. However, in general, regardless of the years of experience, basic cybersecurity awareness training should be provided to every individual. In the later stages of the study, when creating a cyber risk matrix, the lack of cybersecurity awareness training is emphasized as a significant risk.

In this study, it was aimed to learn whether individuals who answered the survey check if HTTPS is used on the websites they visit. In Figure 5.b, individuals received responses indicating that they have been attacked during internet browsing. In this context, when Figure 5.b is examined, the

likelihood of encountering such a graph is high.

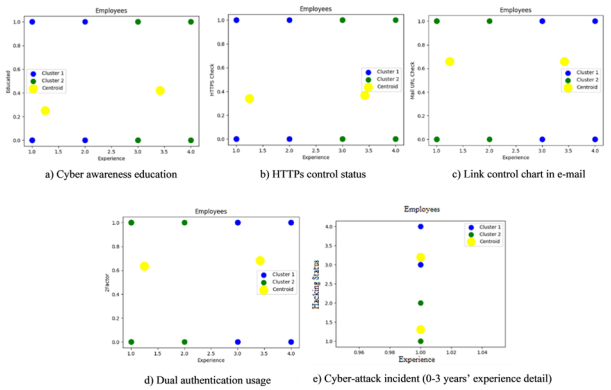

Figure 5. Cyber awareness education graph.

Regardless of experience, it is observed that HTTPS checking is very limited. The responses given by individuals, where 0 means they did not check HTTPS usage, and 1 means they checked it. Observations reveal that the central weight is closer to zero, indicating a deficiency in HTTPS checking among survey participants. In this case, there is a high probability that individuals may enter fake websites on the internet and expose their information to theft.

Recently, the increasing number of people falling into traps confirms this graph. In a news article published in 2023 on the Milli Gazete news site, it was announced that a fake site and application belonging to a popular retail chain were created [6]. The website of this retail chain has been completely copied. Ads and SEO (Search Engine Optimization) adjustments have been made to appear at the top of search engines. When searching for this retail chain on the internet, this fake site appears. Without link and HTTPS checking, there is a high probability of falling into such traps.

Another control method is one of the issues used in Figure 5.c and requires attention. Phishing attacks are a type of cyber-attack that uses disguised email as a weapon [7]. Varieties of phishing attacks use techniques such as text messages, voicemails, or QR (Quick-Response Code) codes. These attacks use social engineering techniques to convince the email recipient that the message is something they want or need (such as a request from a bank).

Referring to a blog post published by Josh F. on the CSO website, it is emphasized, especially for individuals playing a significant role in the company, to check the extensions and links in incoming emails [7]. Many people have fallen victim to phishing attacks that resulted in the theft of their information. Phishing attacks are a matter that companies pay attention to and warn their employees about.

Being cautious and ensuring control in this regard is an expected action from individuals [12].

Incoming links, for example, may come with different domains like g00gle.com instead of google.com. Redirection can also be done through a completely different link. Although a result of around 0.7 is generally obtained for both groups, raising it to 0.8 or even 0.9 levels is necessary for complete security.
The graph regarding the use of two-factor authentication is given in Figure 5.d. Individuals, regardless of their years of experience, mostly actively use two-factor authentication. In this regard, even if individuals' information is stolen, attackers will not be able to gain access unless authorized from their personal devices or applications.

2FA (Two-factor authentication) is an authentication technique that requires users to provide different forms of identification (such as fingerprint verification) and prevents access to their accounts until the password is entered. Using two-step authentication enhances the security of accounts and reduces the likelihood of password theft, decreasing the chances of unauthorized access by attackers. 2FA allows organizations to protect themselves more effectively against phishing attacks and vulnerabilities resulting from human error [8].

2FA can be seen as an additional method that prevents attackers from using stolen information through social engineering, phishing attacks, etc. Hence, the usage of 2FA is crucial, and it proves beneficial in applications, email logins, accounts, etc. [10].

The high usage rate of 2FA is observed due to applications compelling users to use it. Whether using phone applications, SMS (Short Message/Messaging Service), Microsoft Authenticator, etc., even if attackers capture the data, they cannot access users' systems without the code or approval from the users' phones [11].

The next check concerns the question of how often users change their passwords, as shown in Figure 6. When there are more than two answers to this question, representation should be made using Mean Shift. After entering parameters such as age and years of experience, the information about the password change interval, which is the other data to be measured, is used.

In the graph in Figure 6, the X-axis represents years of experience, the Z-axis represents age, and the Y-axis numbers indicate individuals' responses. The responses are sorted as follows: 1 = "Every 1-3 months," 2 = "Every 3-6 months," 3 = "Once a year," 4 = "I don't change it unless required."

Upon examining the graph in Figure 6, it can be observed that individuals with less than 5 years of experience are concentrated in the third option. This implies that individuals mostly change their passwords once a year. Individuals with more than 5 years of experience are seen to change their passwords every 3-6 months, which is a better practice compared to those with less than 5 years of experience. According to research, users changing their passwords every 1-3 months is considered appropriate. However, individuals with less than 5 years of experience changing their passwords only once a year pose a risk for companies.

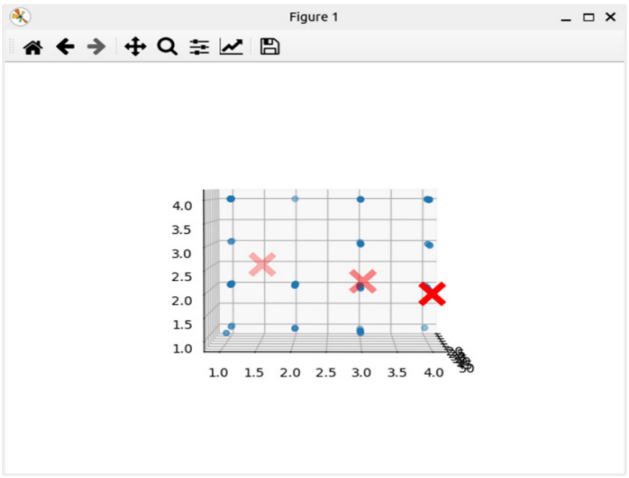

Figure 6. Password change frequency chart.

If individuals use the same password for all their accounts, and if one account is compromised, there is a high probability that other accounts will also be compromised. Each account should have a unique password; for example, the password for Facebook should not be the same as the work password or the mobile banking password [9]. The benefits of changing passwords frequently are as follows:

Prevents continuous access: A hacker may attempt to access your account multiple times within a specific period. Changing your password frequently reduces the risk of the attacker gaining access [9].

Prevents the use of compromised passwords: If you lose or change your devices, someone else might gain access to your passwords. Regularly updating your passwords means that even if an attacker finds an old or compromised password, it will no longer be useful, and your data will be secure [9].

Blocks access obtained by keyloggers: A keylogger is a surveillance technology used to record keystrokes, often used to steal login credentials along with credit card information. Changing your password regularly reduces the likelihood of passwords obtained in this way being useful over any period [9].

Attackers attempting to crack passwords through brute force can easily access user systems when user information is leaked on the internet. Individuals who do not change their passwords frequently are more likely to have their passwords stolen, posing a significant security threat for both users and organizations.

As mentioned at the beginning of this section, all these data and graphs have been examined in a general context so far. All the graphs examined up to this point include a common evaluation of individuals of all experiences and ages. While the used data reflects reality, it can affect each other in detail. For example, looking at Figure 3, it can be seen that the center of gravity of the blue group shifts to the left. In this case, the reason for the shift in the graph is the higher number of individuals with 0-3 years of experience in Group 1. The conclusion to be drawn from this is that individuals with 3-5 years of experience should also be examined separately. The graph in Figure 5.e has been created in detail, focusing only on individuals with 0-3 years of experience. When comparing Figure 5.e with Figure 3, the difference between individuals with 0-3 years and 3-5 years of experience is evident.

As a result, all data and parameters were examined. These reviews are kept on a general and detailed basis in a separate table. A lot of testing and detection has been done. The data obtained from these graphs were compared with each other and connected, and outputs were prepared for the next step, which is to create a risk matrix. These tables and outputs will be discussed in detail in the next section.

GRAPHICS IN DETAIL

The analyzes in the graphics below were used for the detailed part of the matrix. Detailed graphs were created for each question type and the resulting points were shown in the matrix.

In Figure 7, the cyber awareness training status of employees was measured in detail. For employees with 0-3 years of experience, results of 0.2, for employees with 3-5 years of experience, results of 0.4, for employees with 5-10 years of experience, results of 0.5, for employees after 10 years of experience, results of 0.4 were obtained.

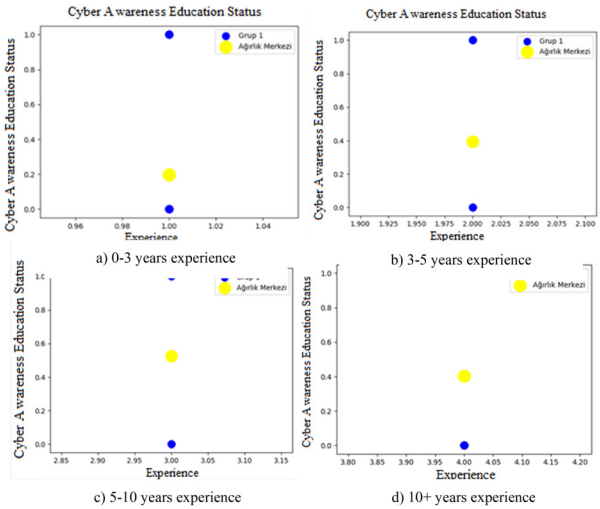

Figure 7. Cyber awareness education status (0-3 years' experience detail chart).

In Figure 8, the HTTPS control status of the employees is measured in detail. For employees with 0-3 years of experience, results of 0.2, for employees with 3-5 years of experience, results of 0.5, for employees with 5-10 years of experience, results of 0.5, for employees after 10 years of experience, results of 0.2 were obtained.

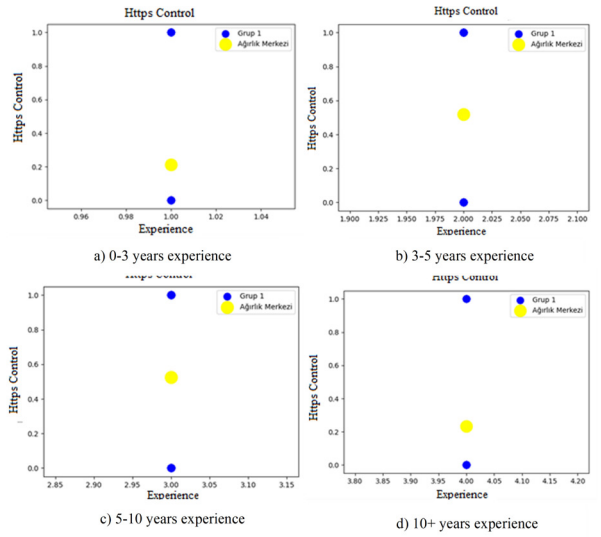

Figure 8. HTTPS check status

In Figure 9, the e-mail url control status of the employees is measured in detail.

For employees with 0-3 years of experience, results of 0.6, for employees with 3-5 years of experience, results of 0.6, for employees with 5-10 years of experience, results of 0.8, for employees after 10 years of experience, results of 0.5 were obtained.

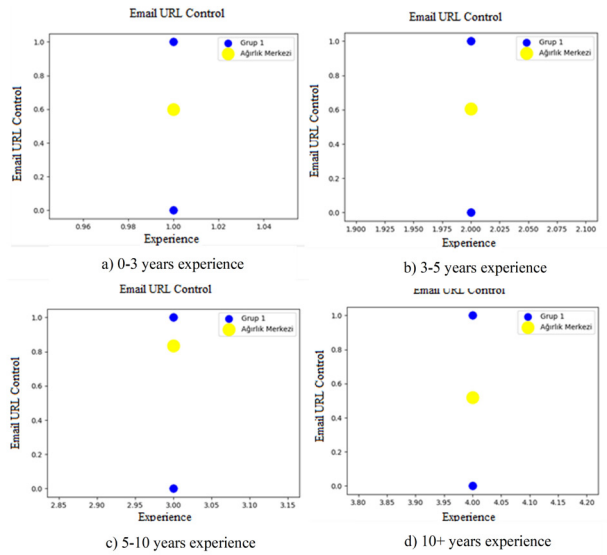

Figure 9. Mail url check status

In Figure 10, employees' 2FA usage is measured in detail. For employees with 0-3 years of experience, results of 0.6, for employees with 3-5 years of experience, results of 0.6, for employees with 5-10 years of experience, results of 0.8, for employees after 10 years of experience, results of 0.8 were obtained.

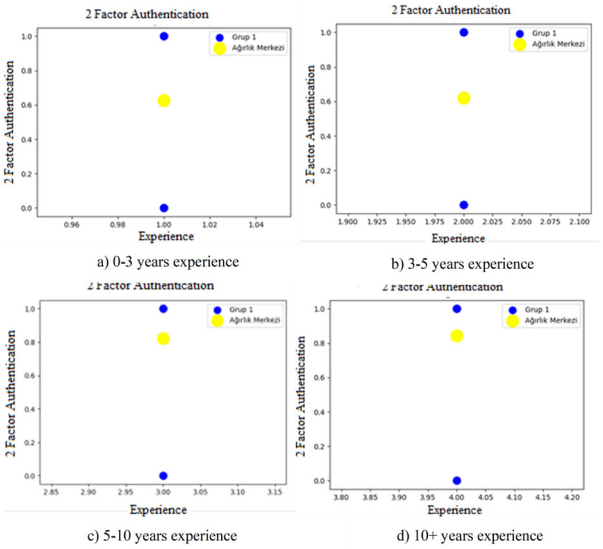

Figure 10. 2FA Usage Case (0-3 years' experience detail chart).

In Figure 11, the cyber incidents experienced by employees are measured in detail. For employees with 0-3 years of experience, results of 1.2 and 3.1, for employees with 3-5 years of experience, results of 1.9 and 3.3, for employees with 5-10 years of experience, results of 2.5 and 6, for employees after 10 years of experience, results of 2.2 and 5.1 when looking at the 2 centroids were obtained.

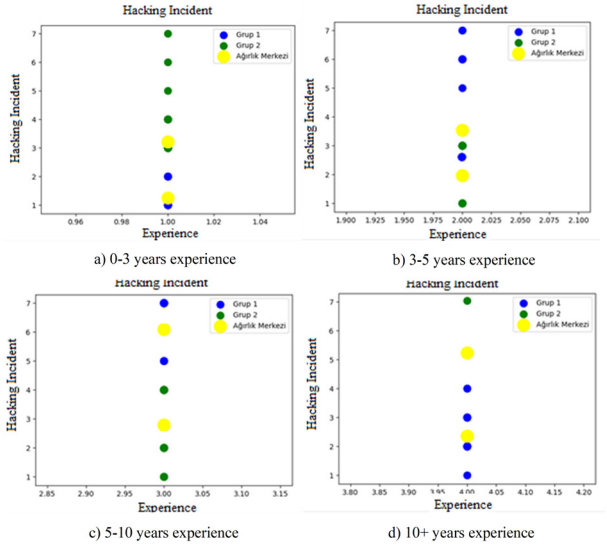

Figure 11. Hacking incident

In Figure 12, the frequency of employees changing their passwords is measured in detail. For employees with 0-3 years of experience, results of 1.7 and 3.8, for employees with 3-5 years of experience, results of 1.5 and 4, for employees with 5-10 years of experience, results of 1.6 and 3.3, for employees after 10 years of experience, results of 1.7 and 3.6 when looking at the 2 centroids were obtained.
These will be used in the risk matrix.

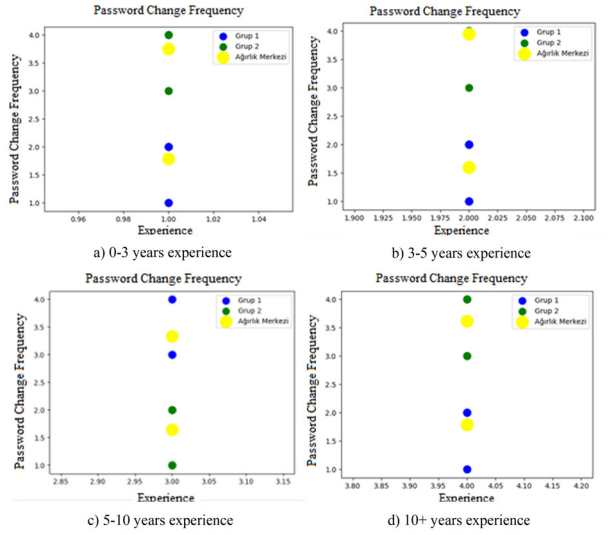

Figure 12. Password change frequency

In Figure 13, employees' password storage method is measured in detail. For employees with 0-3 years of experience, results of 1.3 and 3.2, for employees with 3-5 years of experience, results of 1.8 and 3, for employees with 5-10 years of experience, results of 1.8 and 3.2, for employees after 10 years of experience, results of 1.2 and 3.2 when looking at the 2 centroids were obtained.

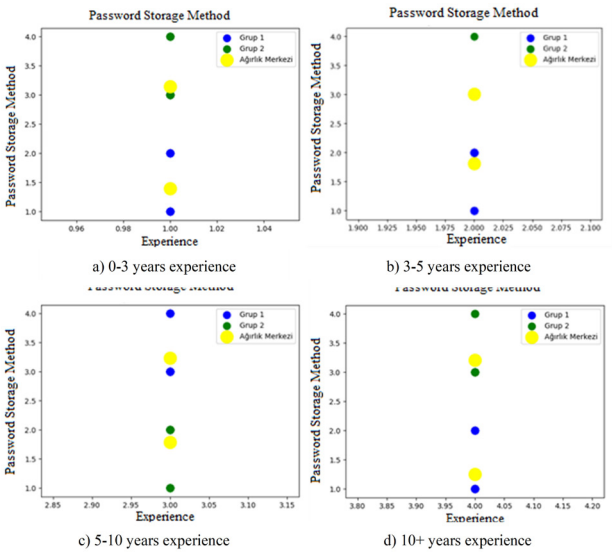

Figure 13. Password storage method

Creation of Risk Matrix and Actions

After the data set consisting of survey data was formatted, the data and algorithms were tested. After this, the algorithms run on real data were tested. Some of the reviews and comments are described in the previous topic. However, all of them were carried out and the results obtained as a result of the algorithm are stored numerically in Table 1. The data collected in this way were compared with each other by looking at the table prepared on a general and detailed basis.

Table 1. General and detailed algorithm results.

	General							
Field/ Year	**0-3**		**3-5**		**5-10**		**10+**	
1	0.3		0.3		0.4		0.4	
2	0.2		0.2		0.4		0.4	
3	0.3		0.3		0.4		0.4	
4	0.7		0.7		0.7		0.7	
5	0.6		0.6		0.7		0.7	
6	4.0		4.0		4.0		6.0	
7	2.5		2.5		2.0		1.5	
8	2.5		2.0		2.5		1.0	
	Detailed							
Field/ Year	**0-3**		**3-5**		**5-10**		**10+**	
1	0.2		0.4		0.5		0.4	
2	0.2		0.5		0.5		0.2	
3	0.6		0.6		0.8		0.5	
4	0.6		0.6		0.8		0.8	
5	0.6		0.8		0.8		0.7	
6	1.2	3.1	1.9	3.3	2.5	6	2.2	5.1
7	1.7	3.8	1.5	4	1.6	3.3	1.7	3.6
8	1.3	3.2	1.8	3	1.8	3.2	1.2	3.2

The fields in Table 1 and their meanings are as follows.

1. Cyber awareness training.
2. HTTPS control.
3. Mail URL control.
4. Use of dual verification.
5. KVKK Information
6. Cyber incident.
7. Password change frequency.
8. Password storage method.

The aim of the study is to interpret the dataset. Based on these interpretations, it seeks to identify under which conditions risks arise and what actions need to be taken. Depending on the company's needs, survey questions and actions may vary. The algorithms can be rerun with modified versions, and new actions and matrices can be determined. However, the established questions have been prepared based on specific research results and with the approval of experts in the field. These questions are sufficient as they are value-adding and universally applicable. Additionally, since the survey participants are not from a single company or profession, the questions are suitable for general use. In short, the conducted study is general, making it effective for any company.

The generated cyber risk matrix differs from matrices that only involve technical information and do not consider human behaviors. Typically, a cyber risk matrix involves a technical examination. In the matrix created for the study, individuals' potential cyber risks based on years of experience are illustrated. For example, it has been observed that individuals with more than 10 years of experience prefer using a notepad as a password storage method. The level of cyber risk posed by these individuals in terms of password storage is

determined to be high.

The data in Table 1 encompasses the results of detailed and general studies conducted as a result of the algorithm. When these questions and answers are examined, it is possible to observe similarities within certain groups. Following this observation, general headings have begun to be created for the cyber risk matrix. In some cases, a title is associated with two questions, while in other cases, it is associated with a single question.

Speaking of titles, the first heading is "Cyber Education." This heading pertains to individuals' basic cybersecurity education status. The second heading is "Cyber Inquiry." This heading includes individuals' HTTPS checks on the sites they visit and the inspection of attachments and links in emails. It indicates whether individuals have the ability to perform cyber inquiries. The third heading is "Password Protection." This heading concerns how individuals store their passwords. The fourth heading is "Password Change." This heading indicates the frequency with which individuals change their passwords and the importance of changing passwords frequently. The fifth heading is "Additional Measures." This heading covers the additional applications, password creation, protection measures, extra plugins, etc., used by individuals for protection in the online environment. The sixth heading is "Legal Authority." This heading encompasses individuals' knowledge of GDPR, procedures to follow in the event of a cyber-attack, and generally their knowledge of legal procedures. After determining the titles and domains of impact, a matrix has been created. Under these headings, individuals were separated by years of experience and added to the matrix. When adding data to the matrix, the structure in Table 2 was created by starting from a less risky level and increasing the risk level downward.

The numbering of the areas determined according to Table 2 is as follows:
1. Cyber awareness training.
2. Cyber interrogation awareness.
3. Password protection.
4. Password change frequency.
5. Additional measures.
6. Legal dominance.

Table 2. Cyber risk matrix.

Field/ Year	0-3 Year	3-5 Year	5-10 Year	10+ Year
6	Low	Low	Low	Low
5	Low	Low	Low	Low
4	Medium	Medium	Low	Low
3	Medium	Medium	Low	High
2	Medium	Medium	Medium	High
1	High	High	High	High

Examining the cyber risk matrix in Table 2, if we focus on the first heading, which is cybersecurity education, it is observed that the risk is high in all individuals regardless of years of experience. Therefore, for the first heading, which is cybersecurity awareness training, it has been indicated that the cyber risk is high regardless of years of experience. In this case, basic cybersecurity awareness training must be provided to everyone entering the company, regardless of years of experience.

Looking again at the cyber risk matrix in Table 2, interpretation has been made for the second heading, which is cyber inquiry. For this heading, when looking at individuals with more than 10 years of experience, the cyber risk is very high. When examining the graph in Figure 10, especially for individuals with more than 10 years of experience, the center of gravity has gathered around answer one. It is observed that these individuals store their passwords in a notepad. Therefore, individuals with more than 10 years of experience will pose a risk to the company. For individuals with other years of experience, the risk is at a moderate level.

Actions that companies should take for new employees or existing employees with more than 10 years of work experience are as follows: These individuals should be recommended password storage applications. The use of these applications can be taught, and the importance of storing passwords can be emphasized. Since this topic may be risky for individuals with more than 10 years of experience, it is necessary to be instructive and guiding to these employees. The same situation applies to individuals with a moderate risk level. However, high priority should be given to employees with more than 10 years of experience or new employees. The titles in the risk matrix and the precautions to be taken for these titles are as follows.

General cyber awareness training should be provided. People are informed about what may happen as a result of cyber-attacks, past individual or corporate cyber events, etc. should be informed about the issues and their importance should be emphasized.

Awareness should be raised among the relevant people with examples of what every action taken in the field of cyber interrogation, a site visited, an e-mail received, a file downloaded from the internet, or a link clicked, can lead to. In addition, they should be informed about how they can check this issue and what they should pay attention to. People should be trained on password protection. People should be informed about what might happen if they fail to keep their passwords well. Encrypted applications used worldwide to store passwords should be mentioned and the use of these applications should be encouraged. Advice should be given about changing existing passwords and switching to these applications.

Regarding changing passwords, people should be explained what might happen if they do not change their passwords. If people do not change their passwords, the methods used by attackers to obtain passwords should be mentioned. Additionally, applications that can be used when creating a password should be shown and their use should be encouraged. The points to be taken into consideration to create the correct password should be shown. Passwords must contain at least 1 numeric

character and 1 special character, and employees must be informed about creating and using complex passwords.

As additional precautions, add ad blockers, HTTPS enforcement, trusted link checking, etc. to the browsers used. Plugins and applications that will perform the operations should be mentioned. The use of these applications and plug-ins should be encouraged within the company.

People should be informed about what they can do legally. Information about KVKK should be given. People's rights should be taught. You should be taught who should be notified of this incident and what procedures should be followed in case of cybercrime occurring inside and outside the company.

CONCLUSION

This study focuses on addressing the human factor, which is a significant vulnerability source in the field of cybersecurity. The goal of the study is to create risk profiles based on the behaviors of employees within the company and develop effective measures accordingly. The findings obtained through the use of machine learning tools provide a valuable resource for strengthening companies' cybersecurity strategies.

The main focus of the study is to categorize employees into specific groups using K-means and Mean Shift algorithms. The aim is to identify similar behaviors within these groups and determine a common action. The risk matrix derived from combining the obtained groups with parameters such as age and experience provides companies with a better understanding of cybersecurity risks, enabling them to develop strategies accordingly.

This study offers an approach that goes beyond the technical aspects of preventive measures in the field of cybersecurity by addressing the human factor. The risk matrix created based on employee profiles provides companies with a clear perspective on potential risks in specific departments or age groups, helping them generate customized solutions.

The methodology presented in this study is applicable to companies, schools, government agencies, and even individual lives. Actions are adaptable based on the needs of legal entities or organizations. Surveys can be re-administered based on specific institutions, and new results can be obtained by running algorithms with this information. This allows for the creation of customized matrices and actions for more effective use.

In conclusion, this study offers a comprehensive approach that not only limits cybersecurity to technical measures but also focuses on employee behaviors. By adopting this methodology, companies can optimize their cybersecurity strategies more effectively and comprehensively. Future research is recommended to further develop this approach by integrating more data sources and exploring new algorithms.

References

1. ENISA. (2020, June 4). Threat landscape 2020: Cyber attacks becoming more sophisticated, targeted, widespread and undetected. European Union Agency for Cybersecurity. Retrieved from https://www.enisa.europa.eu/news/enisa-news/enisa-threat-landscape-2020
2. The Hill. (2020, March 13). FBI sees spike in cyber crime reports during coronavirus pandemic. Retrieved from https://thehill.com/policy/cybersecurity/493198-fbi-sees-spike-in-cyber-crime-reports-during-coronavirus-pandemic/
3. Rajasekharaiah, K. M., Dule, C. S., & Sudarshan, E. (2020, December). Cyber security challenges and its emerging trends on latest technologies. IOP Conference Series: Materials Science and Engineering, 981(2), 022062. https://doi.org/10.1088/1757-899X/981/2/022062
4. Tirumala, S. S., Sarrafzadeh, A., & Pang, P. (2016). A survey on internet usage and cybersecurity awareness in students. In 2016 14th Annual Conference on Privacy, Security and Trust (PST) (pp. 223-228). IEEE. https://doi.org/10.1109/PST.2016.7906931
5. Avcı, Ü., & Oruç, O. (2020). Üniversite öğrencilerinin kişisel siber güvenlik davranışları ve bilgi güvenliği farkındalıklarının incelenmesi. İnönü Üniversitesi Eğitim Fakültesi Dergisi, 21, 284-303. https://doi.org/10.17679/inuefd.526390
6. Yiğit, M., & Seferoglu, S. S. (2019). Öğrencilerin siber güvenlik davranışlarının beş faktör kişilik özellikleri ve çeşitli diğer değişkenlere göre incelenmesi. Mersin Üniversitesi Eğitim Fakültesi Dergisi, 15, 186-215. https://doi.org/10.17860/mersinefd.437610
7. Yetgin, M., & Karakaya, A. (2020). Karabük Üniversitesi çalışanlarına yönelik kişisel siber güvenlik üzerine araştırma. Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 10. https://doi.org/10.47147/ksuiibf.816171
8. Gündüz, M., & Das, R. (2022). Kişisel siber güvenlik yaklaşımlarının değerlendirilmesi. DÜMF Mühendislik Dergisi. https://doi.org/10.24012/dumf.1122997
9. Tokmak, M. (2023). Öğrencilerin siber güvenlik farkındalık düzeylerinin makine öğrenmesi yöntemleri ile belirlenmesi. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 28. https://doi.org/10.53433/yyufbed.1181694
10. Çam, H., Aslay, F., & Özen, Ü. (2019). Yükseköğretim kurumlarında bilgi güvenliği farkındalık düzeylerinin ölçümlenmesi. Yönetim Bilişim Sistemleri Dergisi, 5(2), 1-11.
11. Talesh, S. A. (2018). Data breach, privacy, and cyber insurance: How insurance companies act as "compliance managers" for businesses. Law & Social Inquiry, 43(2), 417-440. https://doi.org/10.1111/lsi.12303
12. Kenny, K. S., Merry, L., Brownbridge, D. A., & Urquía, M. L. (2020). Factors associated with cyber-victimization among immigrants and non-immigrants in Canada: A cross-sectional nationally-representative study. BMC Public Health, 20(1). https://doi.org/10.1186/s12889-020-09492-w
13. Tempestini, G., Rovira, E., Pyke, A., & Nocera, F. D. (2023). The cybersecurity awareness inventory (CAIN): Early phases of development of a tool for assessing cybersecurity knowledge based on the ISO/IEC 27032. Journal of Cybersecurity and Privacy, 3(1), 61-75. https://doi.org/10.3390/jcp3010005
14. Bergh, C. M. M. R. d., & Junger, M. (2018). Victims of cybercrime in Europe: A review of victim surveys. Crime Science, 7(1). https://doi.org/10.1186/s40163-018-0079-3
15. Pramoda, M., Pramoda, S., & Correa, Z. M. O. (2022). Luster regained: A novel cyber incident risk prediction model using

machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 1-19. https://doi.org/10.32628/cseit2283125

16. Nurse, J. R. C., Radanliev, P., Creese, S., & Roure, D. D. (2018). If you can't understand it, you can't properly assess it! The reality of assessing security risks in internet of things systems. Living in the Internet of Things: Cybersecurity of the IoT - 2018 (pp. 1-9). https://doi.org/10.1049/cp.2018.0001
17. Cains, M., Flora, L., Taber, D., King, Z. M., & Henshel, D. S. (2021). Defining cybersecurity and cybersecurity risk within a multidisciplinary context using expert elicitation. Risk Analysis, 42(8), 1643-1669. https://doi.org/10.1111/risa.13687
18. Kumar, A., & Singh, R. (2019). A review of K-means clustering algorithm and its applications. International Journal of Computer Applications, 178(24), 1-5. https://doi.org/10.5120/ijca2019919558
19. Huang, C., & Wang, Y. (2019). A survey on mean shift algorithm and its applications. Journal of Computer Science and Technology, 34(1), 1-20. https://doi.org/10.1007/s11390-019-1906-0
20. Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060. https://doi.org/10.3389/fcomp.2021.563060
21. Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 484–494. https://doi.org/10.1037/0278-7393.14.3.484
22. Aloul, F., Zahidi, S., & El-Hajj, W. (2009). Two-factor authentication using mobile phones. In 2009 IEEE/ACS International Conference on Computer Systems and Applications (pp. 641-644). IEEE. https://doi.org/10.1109/AICSSA.2009.5069395