ORIGINAL RESEARCH

Cyber Incidents Risk Assessments Using Feature Analysis

Benjamin Aziz[1] · Alaa Mohasseb[2]

Received: 20 April 2023 / Accepted: 28 July 2023

Abstract
There are a variety of approaches, methods and techniques that organisations use to manage and contain the risk underlying Cybersecurity incidents throughout their digital and electronic infrastructures. Amongst these are data analysis and data mining techniques, which constitute a big part of the methods applied to data gathered from Cybersecurity incidents. In this study, risk is defined as the product of the probability that incident features will be misjudged and the possible risks for incident responses. We apply our idea to a simple case study involving a dataset of Cyber intrusion incidents in South Korean enterprises. In this paper, we investigate a few problems. First, the prediction of response actions to future incidents involving malware and second, the utilisation of the knowledge of the response actions in guiding analysis to determine the type of malware or the name of the malicious code. Second, a new definition of the probability of risk is based on the precision of the machine learning algorithms. This new definition provides more focus, as it better captures scenarios where response actions are initiated causing resources to be used in cases where a Cyber incident is incorrectly classified as one.

Keywords Cybersecurity · Datasets · Risk analysis · Text mining · Machine learning

Introduction

The significance of the Internet in providing communication infrastructure, data transfer, and services across all domains of life for both private and public sectors cannot be overstated. In the complex landscape of communications enabled by the Internet, Cyber incidents have emerged in recent decades as one of the main sources of risk to organizations, businesses and any solutions with software components. These incidents manipulate, on a daily basis, a plethora of technologies, methods and tools, such as viruses, spyware and spam malware, to bring down businesses, or achieve political or societal goals. The nature of such incidents (and their targets) have become ever more complex over the years, resulting in frequent and substantial financial losses to the global economy as well as reputational damage and potential legal implications. The UK government's Home Office recently released a survey [1], which reported the average cost of a significant Cybersecurity incident in 2021 as ranging from £4200 in small businesses to £19400 per incident in medium to large businesses. Therefore, it is hard to deny that Cybersecurity incidents have become a significant threat to organizations worldwide, a threat that requires attention at all levels of technology.

One approach to the problem is to manage the risk of Cyber incidents effectively. One way to achieve this is through the analysis of any data evidence recorded from such incidents, and using that data to predict information about future incidents [2]. This *data-driven* approach then, which leverages machine learning and data mining techniques, can be used to guide risk calculations and inform the stakeholders. In this context, feature analysis plays a crucial role in identifying the most important features related to an incident, leading to the reduction of the risk of misjudging incident characteristics.

Alaa Mohasseb contributed equally to this work.

This article is part of the topical collection "Advances on Web Information Systems and Technologies" guest edited by Joaquim Filipe, Francisco José Domínguez Mayo and Massimo Marchiori.

✉ Benjamin Aziz
benjamin.aziz@bucks.ac.uk

Alaa Mohasseb
alaa.mohasseb@port.ac.uk

[1] School of Creative and Digital Industries, Buckinghamshire New University, High Wycombe, UK

[2] Present Address: School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK

Published online: 15 November 2023

According to the same report by the UK Home Office [1], organisations are increasingly investing more in information security to mitigate Cyber risks, with up to 41% of the surveyed businesses performing risk management activities. Risk, which is informally defined as anything that negatively impacts an organisation's operations, cannot be completely avoided but can be managed [3]. Data mining and machine learning techniques, which can predict incident characteristics, can play a critical role in managing and mitigating risk and reducing its impact [4–15].

In a previous study [16] conducted by the authors of the current paper, the authors underlined the idea that risk probability can be derived from the accuracy measure of data classification tools [17]. Risk probability was defined as the complement of accuracy, and therefore, it could be combined with meaningful impact to derive risk values in a classical manner. In [16] it was shown how this idea can be used to evaluate the risk for a simple case study of real data representing Cyber intrusion incidents collected from a number of Small and Medium Enterprises (SMEs) in Korea, where text classification tools are trained using the current datasets to predict the values of certain features. This new paper adds to the work of [16] by including a new definition of the probability of risk based on the precision of the machine learning algorithms. This new definition provides for more focus, as it better captures scenarios where response actions are initiated causing resources to be used in cases where a Cyber incident is incorrectly classified as one. In [16], this definition was limited to the accuracy of the algorithm, which is a more generic measure that incorporates some risk-irrelevant scenarios, such as when a "non-incident" is correctly classified as such. As a result, this paper provides a refinement of the original model of risk defined in [16].

The rest of the paper is structured as follows. In "Related Work", we give an overview of related work. In "A Cyber Intrusion Dataset", we give an overview of the Cyber intrusion incidents dataset used in the case study. In "Experimental Study and Results", we discuss the experimental study and the results obtained. In "A Feature Prediction-based Formula for Risk", we introduce our idea that risk can be defined based on the accuracy of the classification algorithms for the class of problems being predicted. In "Risk Analysis", we apply our idea of calculating risk based on prediction accuracy to the case study dataset. Finally, in "Conclusion", we conclude the paper and give directions for future work.

Contributions

The main contribution of this paper is to define a new model of organisational risk that is dependent on the performance of classification algorithms in correctly predicting features related to future Cybersecurity incidents. We outline two kinds of such risk; the first incorporates the Accuracy measure [17] of classification algorithms, whereas the second incorporates the Precision measure [17]. We consider, in both cases, that a wrong prediction will result in acting incorrectly towards future incidents, which may lead to the waste of resources and unnecessary costs.

Related Work

Cybersecurity challenges and threats are prevalent in present-day Internet technologies. As a result, safeguarding the underlying systems and protecting business assets have become crucial aspects for every enterprise. With the increasing prevalence of Cybersecurity threats, organizations are facing vulnerabilities, leading to a surge in research to address these challenges from various perspectives. Several probabilistic and statistical methods for risk assessment have been suggested in the literature, including those discussed in [18–23]. Nonetheless, there has been a recent surge in the use of machine learning for Cybersecurity and risk management due to its superior effectiveness compared to statistical risk models, as demonstrated by the findings in [24]. For example, the use of Naïve Bayes, *k*-Nearest Neighbor, and neural networks have been explored in the context of spam filtering [25], and in [26], the authors demonstrated that the application of the Grey Wolf Optimization algorithm [27] to optimize the parameters of the Support Vector Machines algorithm can lead to a Cybersecurity prediction model with fewer prediction errors.

In [28], the authors proposed MADE (Malicious Activity Detection in Enterprises) to identify malicious activities in enterprise networks and assess the risk of external connections based on predicted probabilities. Meanwhile, [29] explored the combination of supervised and unsupervised learning to capture various Cybersecurity incidents such as malware and malicious emails, using the network structure of dark-web forums data to predict these incidents. In [30], the authors presented a data mining approach to highlight risk factors of network security incidents, leveraging rule mining to detect anomalous patterns and prevent their risk. Similarly, [31] proposed a decision tree-based risk prediction algorithm to minimize the risk of data sharing among financial firms, while [32] introduced a unified risk assessment framework for SCADA networks that incrementally adjusts risk parameters using both historical and real-time observations. Additionally, [33] presented a user-centric machine learning approach for classifying Cybersecurity incidents and categorizing them according to different risk levels. In [34], the authors introduced a new approach to quantifying a company's Cybersecurity risk based on text analytics and advanced autoencoder machine learning techniques. While, [35] proposed a predictive model for risk analysis that calculates risk based on future threat probabilities instead of historical frequencies and surveys conducted by [36, 37] highlight more detailed works related to applications of machine learning techniques to Cybersecurity.

Text mining has also been widely applied to cyber incident detection. In [38], the authors proposed a technique based on analyzing byte n-grams using common N-Gram analysis, which employs profiles to represent classes to identify malicious code. The approach yielded 100% accuracy on the training data and 98% accuracy in a 3-fold cross-validation. Similarly, in [15], the authors presented a method for analyzing suspicious files by extracting OpCode n-gram patterns from the disassembled data of the files to detect unknown malicious code. OpCode n-gram patterns can then be integrated into anti-virus programs as signatures. The evaluation was conducted on a test collection comprising more than 30,000 files, using various settings of OpCode n-gram patterns of different size representations and eight types of classifiers. The results showed the proposed method achieved accuracy higher than 96%.

In the field of telemedicine applications, where malware programs can compromise user privacy, text categorization has been applied as a method, for example, in [39], to identify the characteristics of normal and malicious user behavior by analyzing the data stored in the log files of web servers. On the other hand, [40] proposed a framework for intrusion detection based on system calls, which utilizes text processing and data mining techniques. Suspicious system calls are first analyzed textually and then grouped through the K-means method [41] to identify whether they belong to the group of malicious calls.

In [4], the authors applied data mining and text classification techniques to identify security threats by extracting pertinent information from diverse unstructured log messages. Similarly, the authors in [5] proposed a text mining-based anomaly detection model for detecting HTTP attacks in network traffic, which employs n-gram text categorization and term frequency-inverse document frequency methods.

The authors in [42] presented an approach for automatically detecting malicious code through n-gram analysis. The method utilized selected features based on information gain and employed probabilistic neural networks to build and test the proposed multi-classifier system. The individual classifiers generated classification evidence, which was combined using the Dempster-Shafer combination rules [43, 44] to produce the final classification results for new malicious code. Experimental results demonstrated that the proposed detection engine outperformed the classification results of individual classifiers.

Other studies have employed broader data mining methods that are not restricted to text-based analysis. For instance, the model in [7] utilized hooking techniques to trace the dynamic signatures that malware programs try to conceal. The authors used behavior records to train the classification model and construct a description model. Separately, a method utilizing data mining techniques for detecting spyware was proposed in [13]. The framework utilized a breadth-first search approach, which is effective in detecting viruses and similar software. The accuracy of the method for spyware detection was experimentally determined to be 90.5%. Finally, in [45], the authors presented an integrated architecture to counter surveillance spyware. The architecture employed features derived from both static and dynamic analysis, which were ranked based on their information gains. For each client, a Support Vector Machine classifier was created, and the server gathered reports from all clients to retrain and distribute the new classifier instance to each client. The proposed spyware detection system achieved an overall accuracy rate of 97.9% and 96.4% for known and unknown surveillance spyware programs, respectively.

While these studies offer valuable insights into risk assessment and Cybersecurity analysis, they tend to have a limited scope, as they only consider correctly predicted incidents and ignore the impact of wrongly predicted incidents. Therefore, this paper aims to broaden this scope by defining risk as the product of the probability of *misjudging* incident features and the potential impact of such misjudgment may have on the organization.

A Cyber Intrusion Dataset

The dataset used in our case study represents Cybersecurity intrusion incidents in five SMEs in Korea [46], collected over a period of ten months from 1 January 2017 until 31 October 2017 by the KAITS Industrial Technology Security Hub [47]. As a public-private partnership, the Hub aims to encourage the sharing of knowledge, experience and expertise across Korean SMEs. The data for each SME is stored in a separate file. There is a total of 4643 entries (incidents) in the dataset, divided over five files of a total size of 280KB compressed. The following six features (i.e. metadata), labelled $\ell_1, \ldots, \ell_6$, are included to define the metadata describing these incidents:

- *Date and Time of Occurrence* (ℓ_1): this is a value representing the date and time of the incident's occurrence.
- *End Device* (ℓ_2): this is a value representing the name of the end device affected in the incident.
- *Malicious Code* (ℓ_3): this is a value representing the name of the malicious code detected in the incident.
- *Response* (ℓ_4): this is a value representing the response action that was applied to the malicious code.
- *Type of Malware* (ℓ_5): this is a value representing the type of malware (malicious code) detected in the incident.

- *Detail* (ℓ_6): this is a free text value to describe any other detail about the incident.

An example entry from this dataset looks like the following:

`(14/02/2017 11:58, rc0208-pc, Gen:Variant.Mikey.57034, deleted, virus, C:\Users\RC0208\AppData\ Local\Temp\is-ANFS3.tmp\SetupG.exe)`

We focus next on two of the above features, namely `malicious code` and `response`, as a running example of the application of our model presented here. One can consider *any* combination of these features when formulating the research question. However, our focus here will be the following question:

$\mathfrak{R}$. *Given the current dataset, how can we predict the type of* `Response` *based on the type of* `malicious code`*?*

We shall call $\mathfrak{R}$ our *prediction question*. We can formulate any prediction questions from the above set of features, for example, as was shown in [46]. However, here, we focus on $\mathfrak{R}$ for the rest of the paper. In addition to the above metadata, the dataset also contains statistics on the technical responses to incidents carried out by the five SMEs. This is included in the form of the number of tickets issued in response to the occurring Cyber incidents.

Experimental Study and Results

The objective of the experimental study is to assess the risk calculation of Cyber incidents using feature analysis. Four machine learning algorithms were used for the classification process: J48 Decision tree (J48), RandomForests (RF), Naïve Bayes (NB) and the Support Vector Machine (SVM) algorithm. The data distribution in the KAITS dataset is shown in Table 1.

The experiments were set up using 10-fold cross-validation, and typical performance indicators were used, such as Accuracy, Precision, Recall, and the F-measure [17]. These are defined by the following formulæ:

$$Accuracy = \frac{Number\ of\ correct\ predictions\,(TP + TN)}{Total\ number\ of\ predictions\,(TP + TN + FP + FN)}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where True Positive (TP) is a positive instance classified correctly as positive, True Negative (TN) is a negative instance classified correctly as negative, False Positive (FP) is a negative instance classified wrongly as positive and False Negative (FN) is positive instance classified wrongly as negative. Accuracy, therefore, provides a generic measure of the *goodness* of the classification model. On the other hand, Precision places emphasis on the level of error exhibited by the model in relation to incorrectly identifying positives, whereas Recall places that emphasis on the level of error related to *not* identifying positive cases. Finally, the F measure is a balance between Precision and Recall, albeit avoiding incorporating true negatives, since these are usually of little value to most business cases.

We focus, in what follows, on Accuracy and Precision only, since we are only interested in those cases where the organization responded to the incident, i.e. TP and FP cases (as we shall see later, this means that the organization issued a *ticket* and dispatched resources dealing with the incident), as a measure of our risk probability, as opposed to the cases where an actual incident took place but was missed, i.e. FN cases (therefore eliminating the Recall and F measures), since no form of impact is reported in relation to these cases.

Results

In this section, we present the results of the accuracy and precision of the machine learning algorithms used in our case study for the prediction question $\mathfrak{R}$. The accuracy of these algorithms is summarised by the percentages in Table 2 for each company. The overall results for the identification of the different types of responses based on the given malicious code indicated that SVM had the best accuracy

Table 1 The KAITS dataset data distribution

Company name	Total number of incidents
Company 1 (DF)	932
Company 2 (MT)	633
Company 3 (SE)	923
Company 4 (EP)	448
Company 5 (MS)	1707

Table 2 Accuracy of the classifiers for identifying the types of response based on malicious code—best results are highlighted in bold

Company name	J48	SVM	RF	NB
Company 1 (DF)	83%	**87%**	82%	84%
Company 2 (MT)	86%	**87%**	**87%**	85%
Company 3 (SE)	**89%**	**89%**	**89%**	85%
Company 4 (EP)	86%	**91%**	84%	87%
Company 5 (MS)	**93%**	**93%**	**93%**	89%

for all five companies In addition, most classifiers could not identify response categories such as “none", “blocked" and “deleted".

On the other hand, Table 3 shows the overall results for the precision of these algorithms with respect to each company again when answering the prediction question $\mathfrak{R}$. In this case, we found that the NB classifier had the best precision overall, achieving top precision for three of the five companies.

So far, both of these sets of results provide an indication of the correct prediction performance for each algorithm with respect to the data provided by each company. However, we next look at the reverse side of these results, which provides an indication of the incorrect performance of these algorithms, and hence their riskiness.

A Feature Prediction-based Formula for Risk

We start first by reiterating the classical formula for risk, suggested by IBM’s Robert Courtney, Jr. back in 1977 [48]:

$$risk = probability \times impact$$

as a product of probability and impact. In our case, we assume that impact is defined as separate classes or levels, each representing some qualitative and/or quantitative value concepts. For example, we can express the impact of the leaking of credit card information from a database as a set $\{2M, \}\}legalaction\varepsilon, \}\}reputation\varepsilon\}$, to denote that such leaking would cost the organisation $2M$ as well as impact resulting from legal actions and hits to its reputation. In an impact analysis, an organization could assign any ontological values to this set. More generally, we define $\mathcal{M} = \{m_1, \dots, m_k\}$ as the set of impact levels (qualities or quantities) that an organisation may use to express impact related to that organisation. It is possible to assume further that $\mathcal{M}$ is ordered by some partial ordering relation, $\sqsubseteq_{\mathcal{M}}$, which specifies how some (all) of the levels, $m_1, \dots, m_k$, may compare to one another. For example, $\mathcal{M}$ could refer to some monetary values, such as money, or some computational values such as the increase or decrease in available processing power or time.

Table 3 Precision of the classifiers for identifying the types of response based on malicious code—best results are highlighted in bold

Company name	J48	SVM	RF	NB
Company 1 (DF)	82.8%	77.3%	77.3%	**83.2%**
Company 2 (MT)	58.5%	**59%**	**59%**	57.2%
Company 3 (SE)	58.3%	58.3%	58.3%	**63.2%**
Company 4 (EP)	33.9%	64.4%	49.9%	**64.7%**
Company 5 (MS)	**57.4%**	**57.4%**	**57.4%**	54.2%

We assume that a Cyber incident is described by a set of features (labels), which represent the metadata for that incident. For example, in the dataset we consider here, described in "A Cyber Intrusion Dataset", there are six such features. We refer to these features by the variables $\ell_1, \dots, \ell_k$. The impact of *not* predicting a particular feature of an incident, ℓ, given that all or some of the other features are known, is defined using the following function:

$$impact(\ell) = m_\ell \in \mathcal{M}$$

In other words, $impact(\ell)$ defines the impact on the organisation in case the value of ℓ is predicted incorrectly. For example, if ℓ represents the type of response required, say from knowing the malicious code in the incident, then m_ℓ is the impact on the IT infrastructure or the organisation of misjudging this response.

The probability of making such misjudgment on a feature ℓ is referred to by the value, P_ℓ. We define P_ℓ in two ways: the first is based on the definition of Accuracy (as defined in "Experimental Study and Results"), and the second on the definition of Precision (also as defined in "Experimental Study and Results"). The first definition is given as follows:

$$P^A_\ell = 1 - Accuracy_\ell$$

which states that the risk probability is the complement of Accuracy. This definition expresses risk in a general manner; simply as the general inability of the classification algorithm to predict correctly the Cyber incident feature, ℓ. It is measuring the rate of falsely classifying cases, i.e. FPs and FNs, where we are only interested in FPs, and we consider FNs to be adding an element of *noise* or imprecision to the calculation.

The second definition, however, uses the more focused measure of Precision, as defined by the following:

$$P^P_\ell = 1 - Precision_\ell$$

In this case, we are capturing the risk that the classification algorithm will react incorrectly by predicting that some response value is based on a corresponding malicious code value, which is not the case. This type of risk is aimed at situations where incorrectly predicting responses leads to the waste of resources (e.g. as in the issuing of response tickets and dispatching of resources to counter a non-existent incident). We consider this second definition of the probability of risk as more precise since it is interested in the effect of FPs only, without considering FNs, as in the above first definition.

Depending on which choice we make for the probability of risk, this will determine the meaning and nature of the risk calculated. We can thereby define feature prediction-based risk, resulting from the incorrect prediction of some incident feature ℓ, as in the following equation:

$$risk_{\ell}^{X} = P_{\ell}^{X} \times m_{\ell}$$

where $X \in \{A, P\}$, in our case. We demonstrate next the application of this definition on our Cyber intrusion dataset as described earlier.

Risk Analysis

We explain in the following sections, through the use of a simple example from the KAITS dataset [47], our approach to the calculation of risk within the context of feature prediction in Cyber incidents, and using the definitions we introduced so far in the previous section.

Risk Probability

As we mentioned earlier, our main hypothesis rests on the assumption that the incorrect prediction of an incident's feature represents a risk; due to all the consequences (impact) that will result from such misjudgment. This could include misjudgement that leads to unjustified actions, e.g. dispatching of resources unnecessarily, or misjudgement that leads to inaction, e.g. not dispatching resources where there is a real incident requiring those. In our study, we consider both the prediction accuracy and the prediction precision measures to be relevant as measures of risk probability, since both of these provide an indication of misjudgement when it comes to dispatching of resources unnecessarily.

Based on the values of Table 2, which define the accuracy of answering the prediction question $\mathfrak{R}$, Table 4 presents the accuracy-based risk probability values for each of the classification algorithms and for each of the five companies, for this question. They, therefore, define the value of $P^{A}_{\texttt{Response}}$, in relation to our research quesion.

On the other hand, Table 5 demonstrates the risk probabilities calculated this time using the Precision measure values, as given in Table 3. Therefore, Table 5 is defining the value of $P^{P}_{\texttt{Response}}$.

Having defined the values of $P^{A}_{\texttt{Response}}$ and $P^{P}_{\texttt{Response}}$, we now proceed to define impact.

Table 4 Accuracy-based risk probability of the classifiers for identifying the types of response based on the malicious code

Company name/algorithm	J48	SVM	RF	NB
Company 1 (DF)	17%	13%	18%	16%
Company 2 (MT)	14%	13%	13%	15%
Company 3 (SE)	11%	11%	11%	15%
Company 4 (EP)	14%	9%	16%	13%
Company 5 (MS)	7%	7%	7%	11%

Table 5 Precision-based risk probability of the classifiers for identifying the types of response based on the malicious code

Company name/algorithm	J48	SVM	RF	NB
Company 1 (DF)	17.2%	22.7%	22.7%	16.8%
Company 2 (MT)	41.5%	41%	41%	42.8%
Company 3 (SE)	41.7%	41.7%	41.7%	36.8%
Company 4 (EP)	66.1%	35.6%	50.1%	35.3%
Company 5 (MS)	42.6%	42.6%	42.6%	45.8%

Impact

The KAITS dataset does not include any explicit information about the impact incurred as a result of the incidents. It does however include statistics related to the number of tickets issued in response to incidents at each company, and the number of types of responses that were implemented for those incidents. For our purposes, we shall assume a simple model based on the consideration that the issuing of a ticket incurs some cost, and that this cost can be used as one element of the impact resulting from Cybersecurity incidents. This (form of) impact is then combined with the risk probabilities of the previous section to calculate the overall risk.

We assume that for each company, i, the technical response to an incident (i.e. the response to a ticket issued as a result of an alarm raised on a potential incident) costs, on average, a single *unit* for that company, which we term c_i. This is the average cost per response for servicing an incident at company i. For example, if we consider the recent survey published in [1] as a reference on how much an incident costs, then on average, $c_i = £4200$ for a small company and $c_i = £19400$ for a medium to a large company. Therefore, if a ticket responded to at company i is misjudged (i.e. the type of response is misjudged), then in our model, $impact(\texttt{Response}) = c_i$. For simplicity, we assume that $\mathcal{M} = \{c\}$, meaning that we consider the impact as being measured solely based on the *measure of cost* (e.g. currency).

In our dataset, the number of times tickets issued at each company are responded to, is given in the second column of Table 6. Table 6 also defines, in the third column, the impact factor resulting from Cybersecurity incidents in the form of the total cost each company has incurred, which uses the unit of cost per incident, c_i, at that company.

Unfortunately, the KAITS dataset does not report on the average values of c_i, and hence, these will have to remain

Table 6 Example impact resulting from the incidents

Company name	Number of responses to the issued tickets	Total cost estimated by unit of cost
Company 1 (DF)	3925	$3925 \times c_1$
Company 2 (MT)	13	$13 \times c_2$
Company 3 (SE)	27	$27 \times c_3$
Company 4 (EP)	88	$88 \times c_4$
Company 5 (MS)	19	$19 \times c_5$

Table 8 Risk values of the incorrect Precision-based identification of types of responses based on malicious code using example impact

Company name/algorithm	J48	SVM	RF	NB
Company 1 (DF)	$675.1c_1$	$890.975c_1$	$890.975c_1$	$659.4c_1$
Company 2 (MT)	$5.395c_2$	$5.33c_2$	$5.33c_2$	$5.564c_2$
Company 3 (SE)	$11.259c_3$	$11.259c_3$	$11.259c_3$	$9.936c_3$
Company 4 (EP)	$58.168c_4$	$31.328c_4$	$44.088c_4$	$31.064c_4$
Company 5 (MS)	$8.094c_5$	$8.094c_5$	$8.094c_5$	$8.702c_5$

as variables. Ideally, in future extensions of this study, the impact element need to be detailed to include actual cost and resources spent on responding to Cyber incidents. Such definitions of impact could also include qualitative attributes.

Calculating Risk

Finally, based on the probability of risk and the example impact assumed, we can calculate an overall value for the risk itself. We show this calculation for the two types of risk probabilities we defined in "Risk Probability", namely Accuracy-based and Precision-based risk probabilities.

Table 7 first shows the risk values for each of the five companies associated with the incorrect Accuracy-based prediction of the type of response from the malicious code detected in an incident, based on the example impact given in the previous section. The table thus represents a calculation of $risk^{A}_{\texttt{Response}}$.

These values capture, in some sense, the general inability of the organisation to make correct predictions as risk values. On the other hand, Table 8 shows the risk values associated with the Precision-based prediction this time.

These latter ones reflect a more specific risk associated with the organisation's inability to predict when an action is in fact needed. They, therefore, provide a definition of $risk^{P}_{\texttt{Response}}$.

The interpretation of the data in these tables provides for the view that the incorrect prediction of the type of response to an incident will lead to a misjudgment of the kind or level of service required and therefore will lead to no value in return for the cost in the worst-case scenario. Hence the numbers in the tables represent the worst possible costs of incorrect predictions per algorithm parameterised by each company's currency. These numbers can be interpreted as the limit of the acceptable level when making a Cybersecurity decision in the wrong way. However, the real value underlying these data will be determined by the value of the cost units themselves.

Table 7 Risk values of the incorrect Accuracy-based identification of types of responses based on malicious code using example impact

Company name/algorithm	J48	SVM	RF	NB
Company 1 (DF)	$667.25c_1$	$510.25c_1$	$706.5c_1$	$628c_1$
Company 2 (MT)	$1.82c_2$	$1.69c_2$	$1.69c_2$	$1.95c_2$
Company 3 (SE)	$2.97c_3$	$2.97c_3$	$2.97c_3$	$4.05c_3$
Company 4 (EP)	$12.32c_4$	$7.92c_4$	$14.08c_4$	$11.44c_4$
Company 5 (MS)	$1.33c_5$	$1.33c_5$	$1.33c_5$	$2.09c_5$

Conclusion

In this paper, we have presented an approach for defining risk-based on the probability of inaccurate and imprecise predictions of Cyber incident features, such as the type of response to be given to the malicious code used in the incident, and the potential impact of these predictions in terms of the number of responses served. Our method employs text analysis and classification algorithms on a sample Cyber incidents dataset to illustrate this concept.

This approach has significant implications as it integrates data prediction with the enhancement of an organization's risk analysis. With the increasing popularity of risk analysis frameworks in companies and enterprises of all sizes, our proactive approach can serve as a positive impetus for the development of research in the organizational risk analysis domain. Our future plans revolve around expanding this concept to other domains that also entail a notion of risk, including safety and reliability, leading to more comprehensive risk definitions driven by machine learning. We have already demonstrated in [49] how risk can be defined and calculated based on the analysis of large and open datasets for the case of user logins in shared computing environments.

One of the main drawbacks of our study is the missing information on impact and its cost. We aim to expand our study in the future to collect datasets that include details about the impact of Cyber incidents, in particular, information on the quantity and quality of resources used. Such missing information currently restricts the scope of the study to a theoretical perception of risk, and future research will focus on obtaining and generating data that contains both probability and impact values, to gain a concrete understanding of the applicability of our underlying risk model. Additionally, we hope that in future studies, we will aim to expand the scale of the study by applying our analysis to larger Cybersecurity datasets, particularly those available on open platforms such as VCDB [50], SecRepo [51] and CAIDA [52], which include different information and hence pause different research questions. Finally, we aim to investigate the use of other machine learning performance measures, for example, Recall and the *F* measure, to expand our risk definition.

Author Contributions All authors contributed equally to the work presented in the paper.

Funding Not applicable.

Data Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. HM Government: Cyber Security Breaches Survey 2022. https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2022/cyber-security-breaches-survey-2022 2022. Accessed 29 Sept 2023.
2. Martínez Torres J, Iglesias Comesaña C, García-Nieto PJ. Machine learning techniques applied to cybersecurity. Int J Mach Learn Cybern. 2019;10:2823–36.
3. Kaplan S, Garrick BJ. On the quantitative definition of risk. Risk Anal. 1981;1(1):11–27.
4. Suh-Lee C, Jo JY, Kim Y. Text mining for security threat detection discovering hidden information in unstructured log messages. In: Communications and Network Security (CNS), 2016 IEEE Conference On, 2016;252–260. IEEE
5. Kakavand M, Mustapha N, Mustapha A, Abdullah MT. A text mining-based anomaly detection model in network security. Glob J Comput Sci Technol. 2015;14(1):22–31.
6. Norouzi M, Souri A, Samad Zamini M. A data mining classification approach for behavioral malware detection. J Comput Netw Commun. 2016;2016:1.
7. Fan CI, Hsiao HW, Chou CH, Tseng YF. Malware detection systems based on api log data mining. In: Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual, vol. 3, 2015;255–260. IEEE
8. Hellal A, Romdhane LB. Minimal contrast frequent pattern mining for malware detection. Comput Secur. 2016;62:19–32.
9. Lu Y-B, Din S-C, Zheng C-F, Gao B-J. Using multi-feature and classifier ensembles to improve malware detection. J CCIT. 2010;39(2):57–72.
10. Fan Y, Ye Y, Chen L. Malicious sequential pattern mining for automatic malware detection. Expert Syst Appl. 2016;52:16–25.
11. Rieck K, Trinius P, Willems C, Holz T. Automatic analysis of malware behavior using machine learning. J Comput Secur. 2011;19(4):639–68.
12. Ding Y, Yuan X, Tang K, Xiao X, Zhang Y. A fast malware detection algorithm based on objective-oriented association mining. Comput Secur. 2013;39:315–24.
13. Bahraminikoo P, Yeganeh M, Babu G. Utilization data mining to detect spyware. IOSR J Comput Eng (IOSRJCE). 2012;4(3):01–4.
14. Schultz MG, Eskin E, Zadok F, Stolfo SJ. Data mining methods for detection of new malicious executables. In: Security and Privacy, 2001. S &P 2001. Proceedings. 2001 IEEE Symposium On, 2001;38–49. IEEE
15. Shabtai A, Moskovitch R, Feher C, Dolev S, Elovici Y. Detecting unknown malicious code by applying classification techniques on opcode patterns. Secur Inform. 2012;1(1):1.
16. Aziz, B., Mohasseb, A.: Using feature analysis to guide risk calculations of cyber incidents. In: 18th International Conference on Web Information Systems and Technologies. 2022. SciTePress
17. Chinchor N. Muc-4 evaluation metrics. In: Proceedings of the 4th Conference on Message Understanding. MUC4 '92, 1992;22–29. Association for Computational Linguistics, Stroudsburg, PA, USA
18. Sommestad T, Ekstedt M, Johnson P. A probabilistic relational model for security risk analysis. Comput Secur. 2010;29(6):659–79.
19. Shin J, Son H, Heo G. Cyber security risk analysis model composed with activity-quality and architecture model. In: International Conference on Computer, Networks and Communication Engineering (ICCNCE 2013). 2013. Atlantis Press
20. Cherdantseva Y, Burnap P, Blyth A, Eden P, Jones K, Soulsby H, Stoddart K. A review of cyber security risk assessment methods for Scada systems. Comput Secur. 2016;56:1–27.
21. Ruan K. Introducing cybernomics: a unifying economic framework for measuring cyber risk. Comput Secur. 2017;65:77–89.
22. Paté-Cornell M-E, Kuypers M, Smith M, Keller P. Cyber risk management for critical infrastructure: a risk analysis model and three case studies. Risk Anal. 2018;38(2):226–41.
23. Santini, P., Gottardi, G., Baldi, M., Chiaraluce, F.: A data-driven approach to cyber risk assessment. Security and Communication Networks 2019. 2019.
24. Kakushadze Z, Yu W. Machine learning risk models. J Risk Control. 2019;6(1):37–64.
25. Mohasseb A, Aziz B, Kanavos A. SMS Spam Identification and Risk Assessment Evaluations. In: Proceedings of the 16th International Conference on Web Information Systems and Technologies - Volume 1: DMMLACS,, 2020;417–424. SciTePress. INSTICC
26. Lu H, Zhang G, Shen Y. Cyber security situation prediction model based on gwo-svm. In: International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, 2019;162–171. Springer.
27. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw. 2014;69:46–61.
28. Oprea A, Li Z, Norris R, Bowers K. Made: Security analytics for enterprise threat detection. In: Proceedings of the 34th Annual Computer Security Applications Conference, 2018;124–136. ACM
29. Sarkar S, Almukaynizi M, Shakarian J, Shakarian P. Mining user interaction patterns in the darkweb to predict enterprise cyber incidents. Soc Netw Anal Min. 2019;9(1):57.

30. Gounder MP, Nahar J. Practicality of data mining for proficient network security management. In: 2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), 2018; 149–155. IEEE
31. Gai K, Qiu M, Elnagdy SA. Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data. In: 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016;197–202. IEEE
32. Huang K, Zhou C, Tian YC, Tu W, Peng Y. Application of bayesian network to data-driven cyber-security risk assessment in scada networks. In: 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), 2017;1–6. IEEE
33. Feng C, Wu S, Liu N. A user-centric machine learning framework for cyber security operations center. In: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), 2017;173–175. IEEE.
34. Cheong A, Cho S, No WG, Vasarhelyi MA. If you cannot measure it, you cannot manage it: Assessing the quality of cybersecurity risk disclosure through textual imagification. 2019. SSRN
35. Figueira PT, Bravo CL, López JLR. Improving information security risk analysis by including threat-occurrence predictive models. Comput Secur. 2020;88: 101609.
36. Rawat DB, Doku R, Garuba M. Cybersecurity in big data era: From securing big data to data-driven security. IEEE Transactions on Services Computing. 2019.
37. Torres JM, Comesaña CI, García-Nieto PJ. Machine learning techniques applied to cybersecurity. Int J Mach Learn Cybern. 2019;10:1–14.
38. Abou-Assaleh T, Cercone N, Keselj V, Sweidan R. N-gram-based detection of new malicious code. In: Computer Software and Applications Conference, 2004. COMPSAC 2004. Proceedings of the 28th Annual International, vol. 2, 2004;41–42. IEEE
39. Adeva JJG, Atxa JMP. Intrusion detection in web applications using text mining. Eng Appl Artif Intell. 2007;20(4):555–66.
40. Kumar GR, Mangathayaru N, Narasimha G. An approach for intrusion detection using text mining techniques. In: Proceedings of the The International Conference on Engineering & MIS 2015. ICEMIS '15, 2015;63–1636. ACM, New York, NY, USA.
41. Macqueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967;281–297.
42. Zhang B, Yin J, Hao J, Zhang D, Wang S. Malicious codes detection based on ensemble learning. In: International Conference on Autonomic and Trusted Computing, 2007;468–477. Springer.
43. Dempster AP. Upper and lower probabilities induced by a multivalued mapping. Ann Math Statist. 1967;38(2):325–39. https://doi.org/10.1214/aoms/1177698950.
44. Shafer G. A mathematical theory of evidence. Princeton: Princeton University Press; 1976.
45. Wang TY, Horng SJ, Su MY, Wu CH, Wang PC, Su WZ. A surveillance spyware detection system based on data mining methods. In: Evolutionary Computation, 2006. CEC 2006. IEEE Congress On, 2006;3236–3241. IEEE
46. Mohasseb A, Aziz B, Jung J, Lee J. Cyber security incidents analysis and classification in a case study of Korean enterprises. Knowl Inf Syst. 2020;62:2917–35.
47. KAITS: Industrial Technology Security Hub. https://www.kaits.or.kr. Accessed 29 Sept 2023.
48. Robert H. Courtney J. Security Risk Assessment in Electronic Data Processing Systems. In: Proceedings of the June 13-16, 1977, National Computer Conference. AFIPS '77, 1977;97–104. ACM, New York, NY, USA.
49. Aziz B. Analysing potential data security losses in organisations based on subsequent users logins. PLOS One. 2023. https://doi.org/10.1371/journal.pone.0286856.
50. VERIZON: The VERIS Community Database https://github.com/vz-risk/VCDB. Accessed 29 Sept 2023.
51. Mike Sconzo: SecRepo.com—Samples of Security Related Data. http://www.secrepo.com. Accessed 29 Sept 2023.
52. Center for Applied Internet Data Analysis: CAIDA Data. https://www.caida.org/data/overview/. Accessed 29 Sept 2023.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.