SecRiskAI: a Machine Learning-Based Approach for Cybersecurity Risk Prediction in Businesses

Muriel F. Franco[1], Erion Sula[1], Alberto Huertas[1], Eder J. Scheid[1], Lisandro Z. Granville[2], Burkhard Stiller[1]
[1]*Communication Systems Group CSG, Department of Informatics IfI, University of Zürich UZH*
Binzmühlestrasse 14, CH—8050 Zürich, Switzerland
[2]*Computer Networks Group, Institute of Informatics, Federal University of Rio Grande do Sul UFRGS*
Av. Bento Gonçalves, 9500, Porto Alegre, Brazil
E-mail: [franco, huertas, scheid, stiller]@ifi.uzh.ch[1], erion.sula@uzh.ch[1], granville@inf.ufrgs.br[2]

***Abstract*—Cyberattacks have increased in number and severity, negatively impacting businesses and their services. As such, cybersecurity can no longer be seen just as a technological issue, but it must also be recognized as critical to the economy and society. Current solutions struggle to find indicators of unpredictable risks, limiting their ability to perform accurate risk assessments. This work thus introduces SecRiskAI, an approach that employs Machine Learning (ML) to assess and predict how exposed a business is to cybersecurity risks. For this purpose, four ML algorithms were implemented, trained, and evaluated using synthetic datasets representing characteristics of different sizes of businesses (e.g., number of employees, business sector, and known vulnerabilities). Moreover, a Web-based user interface is provided to simplify the risk prediction workflow. The quantitative evaluation performed on SecRiskAI shows a minimal performance overhead and the high accuracy of the ML models, while a case study assesses the feasibility of the overall process for decision-makers.**

I. Introduction

Cyberattacks are a rising threat for governments and companies. As businesses become more digital, they are exposed to an increasing number of threats [4]. Thus, beyond compromising companies and their customers' security and privacy, cyberattacks can negatively impact the economy of businesses and services supported by digital systems [16]. Predictions from Cybersecurity Ventures, a world's leading researcher for the global cyber economy, indicate that cybercrime damages will hit US$ 10 trillion annually by 2025 [13]. It suggests that cybersecurity can no longer be seen just as a technology issue but must also be watched from an economic optic.

There exist several attacks used against general businesses, such as Distributed Denial-of-Service (DDoS), Ransomware, and Phishing. However, some sectors have been the target of specific threats. For example, businesses that handle critical services and information (*e.g.*, hospitals, universities, and finance) tend to be more targeted by ransomware attacks, which encrypt the data to make all systems unavailable. Because of the critical nature of their services and the technical complexities to restore from the attack (requiring days or even weeks), these businesses tend to pay for the rescue asked by attackers. Although there are businesses more targeted for specific attacks, in general, attackers tend not to spend too much time focusing on a particular business but on exploring vulnerabilities in any business they see as potential victims. This happens especially in the case of Small and Medium-sized Enterprises (SMEs), which are the focus of general attacks (*i.e.*, not tailored for a specific company). because attackers know that SMEs usually have lack of in-house cybersecurity expertise.

Therefore, it is important to understand the likelihood of these threats and the risks, such as the economic impacts and business disruption due to cyberattacks. Despite the several risk assessment standards (*e.g.*, ISO 31000 and NIST SP 800-30) and specific frameworks [16], [14] available, organizations still find this activity challenging and are often confronted with a massive volume of unstructured data, which hinders the identification of risks. In this case, traditional techniques may not provide valuable insights and cannot perform an adequate risk assessment due to the amount of data to be processed.

Several studies on possible applications of Machine Learning (ML) algorithms [17], a branch of Artificial Intelligence (AI), have highlighted their ability to process large amounts of structured/unstructured data, extract valuable patterns, learn from historically collected records, and make accurate predictions. Given the characteristics of learning and identifying patterns, ML-based systems can be an ally for the qualitative analysis of potential risks and threats within a company, thus helping in risk assessment and planning of cybersecurity. For example, ML algorithms can be used to correlate specific characteristics and information (*e.g.*, revenue, sector, cybersecurity strategies, and infrastructure) of a company to associate it with a higher or lower risk to have a breach in its cybersecurity.

This work introduces *SecRiskAI* to address the lack of solutions for risk assessment and predicting threats in a straightforward and simplified way. *SecRiskAI* implements four ML algorithms for risk assessment and builds models to predict general and specific threats (*e.g.*, DDoS or Phishing attack). Relevant information and features for the ML-based risk assessment are also presented and described. A Web-based interface is also provided to simplify the process of understanding the business risks in a user-friendly and intuitive way. *SecRiskAI* is evaluated in a *(a)* quantitative way to measure the performance and accuracy of the ML algorithms, and *(b)* qualitative way based on a case study in a selected scenario. The source-code of *SecRiskAI* is publicly available at [3].

The remainder of this paper is organized as follows. Section II reviews related work. Section III introduces *SecRiskAI*, while Section IV contains the evaluation, followed by conclusions and future work in Section V.

II. Related Work

The term risk is generally used to indicate a possibility of loss and/or damage. It usually involves some degree of uncertainty, and the resulting outcome is challenging to predict. Depending on the context, various types of risks can be found, such as business risk, economic risk, and safety risk. In the context of this paper, risk is defined as the probability of a threat happening that can cause economic and reputation losses for a business.

In addition to frameworks from standardization institutes (*e.g.*, NIST Risk Management Framework, ISO 31000, and TOGAF Security Guide), different approaches have been proposed to address the challenges of cybersecurity risk assessment [16], [12], [14]. Specific models have been proposed along the years for threat modeling and risk assessment in different scenarios and applications. For example, while NIST guidelines focus on the overall risks of an organization, STRIDE, LINDDUN, and DREAD map each specific type of threat as well as their mitigation actions.

The state-of-the-art yet shows that there is still room to improve and evolve traditional risk assessment processes by employing novel technologies, such as by exploiting different branches of AI, *e.g.*, ML and Deep Learning (DL). This opportunity emerges by the nature of the risk assessment problem (*e.g.*, nondeterministic attributes to identify risks and a large number of statistical attributes required) and also by the lack of approaches that explore AI to understand risks based on general and specific information available about a company and its systems [1].

Researchers from the National Institute of Standards and Technology (NIST) proposed a model to assess cybersecurity risks to support investments strategies in network security [12]. The work highlights how ML can be used as a foundation for cybersecurity investments in different scenarios, *e.g.*, those that use remote work tools, Internet of Things (IoT) devices, and mobile elements. In the field of cybersecurity, research tended to focus mainly on leveraging ML to detect various types of cyberattacks and recognize breaches [17]. Examples include the identification of malicious traffic [7], anomaly detection [18], and attack mitigation [10]. Besides, ML has the potential to change the cybersecurity and risk assessment landscape significantly. However, there are still few efforts specifically focusing on cybersecurity risk assessment.

An ML-based model for cybersecurity risk assessment in smart cities has been proposed [8]. The novelty here relies on the use of Artificial Neural Network (ANN) to process massive security datasets to provide faster response times in critical situations. In simulations composed of 10,000 vectors with 38 features each (*e.g.*, device type, probability of specific attacks, and device cost), the accuracy of the model was 97%. The authors argued that ML techniques might have a key role in risk assessment in environments where considerable amount of data and hidden dependencies are present. In the automotive industry, in [2], the authors focused on investigating possible factors that have the most significant impact on car accidents. The authors specifically applied Decision Tree (DT) and ANN to identify relevant patterns and detect the most frequent key factors involved in car accidents. Similarly, [19] applied different ML algorithms to classify the risk of severe injury based on over 5,000 traffic accident records.

Solutions have also been proposed to assess risks in the Energy & Nuclear sectors. For nuclear energy, [15] relies on a Support-Vector Machine (SVM) classifier to detect anomalies and evaluate the risks of possible malfunctions of components. In another work, [11] proposed a DL-based model to assess economic risks in virtual power plants. The authors exploited two techniques called Naive Bayes and the J49 bagging tree model. The initial results suggested a promising path for AI as an ally for measuring and understanding economic impacts during cybersecurity planning. However, this study and other related work provide evidence that challenges still have to be addressed in the AI field, especially those related to the lack of explainability of DL algorithms and sufficient cybersecurity information sharing to train these algorithms.

In [21], a fuzzy probability Bayesian network is proposed for dynamic risk assessment in industrial control systems. The solution considered four case studies based on a chemical reactor control system under attack. The authors stated that risk assessment approaches have to consider the specific functionality and features of the analyzed scenarios. In another work, [22] designed a multi-model incident prediction and risk assessment approach for industrial control systems. The approach can use multiple models to predict the impact of cyberattacks and assess the risk of unknown attacks. A multilevel Bayesian network is also developed, composed of an attack model, a function model, and an incident model to describe risk propagation due to cyberattacks.

Based on an extensive literature review, four main ML techniques (*cf.* Section III-B) were selected to be implemented within the context of *SecRiskAI* due to their identified potential (*i.e.*, characteristics, datasets required, and performance) for the scenario explored in this paper. Thus, by implementing these techniques, *SecRiskAI* can help businesses that do not have the technical expertise to conduct complex risk assessments or even use them as an initial cybersecurity planning step in sectors where information are hard to obtain, such as businesses that do not have a well-defined strategy for monitoring and managing their data, systems, and resources.

III. *SecRiskAI* Approach

SecRiskAI approach focuses on predicting the risks of companies to support the planning and deployment of effective cybersecurity strategies, which can avoid technical problems and reduce potential financial losses resulting from cyberattacks. For a qualitative cybersecurity risk assessment and understanding the likelihood of attacks in companies, *SecRiskAI* provides an approach based on three main steps: *(a)* datasets definition and data generation, *(b)* ML model creation, and *(c)* the risk prediction. For that, different data sources and features are determined, and ML algorithms are explored to be employed in different situations and constraints (*e.g.*, limited amount of information and size of datasets available). Also, it is critical for companies, especially those without in-house expertise (*e.g.*, SMEs and micro enterprises), to have an easy and straightforward approach. Therefore, *SecRiskAI* integrates the whole process of risk prediction in a user-friendly and intuitive Web-based Interface. The source-code of *SecRiskAI* and a fully operational prototype is publicly available at [3], including all components, training datasets, and models.

Figure 1 introduces the *SecRiskAI* architecture and stakeholders. First, the user (*i.e.*, companies and decision-makers) accesses the dashboard. The Web-based Interface is designed to provide visibility of business-related risk indicators and, at the same time, increase productivity and better forecasting of important aspects related to the business security. Moreover, through the Web-based Interface, the user can change contextual information (*cf.* Section IV-B).

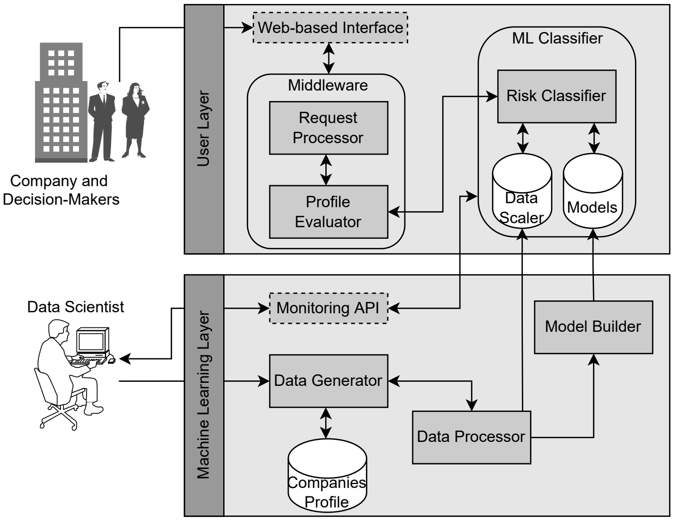


Fig. 1: *SecRiskAI*'s Architecture Overview

The task of using the data provided by the user (*i.e.*, business profile and characteristics) to make risk predictions is performed by the Middleware module. As soon as the request sent by the Web-based Interface is received, the *Request Processor* handles it and forwards the information to the *Profile Evaluator*, which is in charge of contacting the ML models and evaluating the prediction response. The risk prediction starts with a request to the *Risk Classifier*, which is a prediction service included in the ML Classifier module and is essentially used to expose the trained ML models through an Application Programming Interface (API). Additionally, the ML Classifier module also stores the trained ML models, as well as the *Data Scalers*, used to normalize the input data and increase prediction accuracy.

The process of training and validating the ML models occurs in the ML Layer and is usually carried out by data scientists/experts with knowledge about the business and respective sector. This can be also part of consultancy services provided by third-parties. The *Data Generator* component is used to initialize the synthetic data generation process. This data generation is an Python script implemented to generate synthetic labeled datasets based on characteristics of businesses (*cf.* Table I), according to the requirements of the ML algorithms. Next, the data is processed (*i.e.*, Data Processor) and used by the *Model Builder* for training, validating, testing, and building the models. Each action performed by the components of the ML Layer is described in detail in Section III-A. Lastly, a Monitoring API is available for checking the status of the deployed models, retrieving model-specific metadata (*e.g.*, version, creation time, and accuracy) and other metrics about the prediction service (*e.g.*, request duration in seconds and status).

A. *Risk Assessment and ML Workflow*

Once the opportunities of applying ML to cybersecurity risk assessment are defined and well-understood, the process of designing and developing an ML workflow starts. The most critical stage is data collection/gathering. Usually, data is collected from sensors or other different sources and stored for further processing in this phase. However, in the field of cybersecurity risk assessment, companies either do not disclose any information at all or, in some cases, publish various reports that are often incomplete and difficult to extract meaningful results from. Hence, a synthetic data generator approach was designed and implemented to overcome this limitation and feed *SecRiskAI* with data to be used for the training process of the algorithms.

An exploratory analysis was conducted to determine relevant parameters that can increase or decrease the risk of a company. This analysis consists of three main sources: *(a)* public reports from different agencies and companies, such as those from European Union Agency for Cybersecurity (ENISA) and European Digital SME Alliance, *(b)* scientific works indexed by well-known digital repositories (*e.g.*, IEEEXplore, ACM Digital Library, and Google Scholar) that covers the likelihood, severity, and effects of cybersecurity issues in SMEs mostly, and *(c)* interviews with cybersecurity experts and SMEs owners to understand their reality and information asymmetry challenges. It is important to note that this is not an exhaustive analysis but gives indications of the most common characteristics of companies that can be related to the risks of being affected by a cyberattack. Also, the cyberattacks investigated are restricted to Phishing, Ransomware, and DDoS attacks. After such an exploratory analysis of different cyberattacks and corresponding companies' contextual information, the following parameters to be used as a basis for this work were identified:

- **Revenue.** Income generated from normal business activities and operations, and in most cases is also used to classify businesses by providing a scale for determining their sizes.
- **Cybersecurity Investments.** Businesses may have cybersecurity investments strategies in place to ensure a proper

level of defense. This kind of information needs to be taken into consideration during the cybersecurity risk assessment, as it may have an impact on the likelihood of being targeted by a cyberattack.

- **Number of Employees and Training Level.** Information regarding the actual number of employees in a company as well as the corresponding cybersecurity training level (*e.g.*, cybersecurity basic knowledge and phishing training) represent essential contextual information required for assessing possible cyber-risks. The employee training level is measured as *Low*, *Medium* and *High*.
- **Successful/Failed Cyberattacks.** This parameter indicates the number of cyberattacks that the company has already experienced. This includes different attacks (*e.g.*, DDoS and Phishing) that have targeted the organization's infrastructure and resulting in either a financial loss or reputation damage. Failed attempts are also taken into consideration.
- **Known Vulnerabilities.** For an effective and comprehensive risk assessment, it is essential to report any known vulnerabilities of the infrastructure. Vulnerability management is usually a key responsibility of the companies' IT security team. This phase usually involves assessing and reporting any security vulnerability present in the organization's systems. There are a variety of comprehensive tools used for vulnerability scanning, such as nmap, Metasploit, and OWASP. Currently, the total number of known vulnerabilities is defined during the synthetic generation process.
- **External Cybersecurity Advisor.** To further strengthen their cyber resilience (*i.e.*, the ability to prepare for, respond to, and recover from cyberattacks), businesses are encouraged to hire external Cybersecurity Advisor (CSA). During the synthetic data generation phase, a binary value will be generated (either *Yes* or *No*).
- **Risk.** Represents the value of the qualitative risk assessment based on the previously generated parameters. Since the synthetic data generation process is designed to generate historical records of companies operating in comparable industries, the value of the risk column may be derived from past formal or tailored qualitative risk assessment techniques. The generated risk can assume one of the following values: *Low*, *Medium* and *High*.

To generate the information mentioned above, some assumptions were made. First, upper/lower boundaries for each column were specified so that each generated value would effectively lie in the defined range. Table I provides an overview of the determined boundaries as well as examples of values for each generated information. These attributes are also used as input to map the risks according to what is proposed by *SecRiskAI* in Equation 1.

Not all of this information must be available within the company, especially considering SMEs that do not have in-house expertise. This is highlighted in the last column of Table I. Therefore, the Failed Attacks and Known Vulnerabilities are optional for the *SecRiskAI*. Although it is essential to know these metrics for an accurate risk assessment, it is possible to address the lack of this information by understanding the correlation between successful attacks and other statistics available (*e.g.*, economic losses, number of attacks per sector, and trends) from companies from the same sector. This can be adjusted by adding, in the training dataset, labeled data that represents this behavior or trend.

$$
\begin{aligned}
& i_r = \frac{invested\_amount}{business\_value} \\
& e = \frac{nr\_employees}{tot\_empl} * map(employees\_training) \\
& att_r = \frac{succ\_attacks}{max\_attacks} \\
& v_r = \frac{known\_vuln}{max\_known\_vuln} \\
& adv_i = map(external\_adv) \\
& map(x) = \begin{cases} 0, & \text{if } x = Low \\ 1, & \text{if } x = Medium \\ 2, & \text{if } x = High \end{cases}
\end{aligned} \tag{1}
$$

$$
computed\_risk = i_r + e + adv_i - att_r - v_r \tag{2}
$$

It is important to notice that the risk is not randomly generated; rather, it is computed based on the generated attributes shown in Table I using the generalized Equation 2. For the supervised learning process, the dataset must be labeled. As a result, the *computed_risk* output is mapped to either a Low, Medium, or High class. A manual labeling process would be too expensive, since the generated dataset would include thousands of records. Therefore, based on the numeric value of *computed_risk*, a mapping range is defined. This means that each *computed_risk* value is labeled using the range as specified in the end of Equation 1.

TABLE I: Overview of the Dataset Generated Attributes

Information	ID	Range	Priority
Revenue	business_value	0 to 5,000,000	Required
Cybersecurity Investment	invested_amount	0-30% * Revenue	Required
Successful Attacks	succ_attack	0 to 50	Required
Failed Attacks	fail_attack	0 to 50	Optional
Number of Employees	nr_employees	30 to 10,000	Required
Employee Training	employees_training	Low, Medium or High	Required
Known Vulnerabilities	known_vuln	0 to 10	Optional
External Cybersecurity Advisor	external_adv	Yes or No	Required
Risk	risk	Low, Medium or High	-

Figure 2 summarizes the ML workflow implemented by *SecRiskAI*. This is well-known ML workflow and has been adapted to fulfill the demands of the approach, such as the data processing techniques and the decision on ML algorithms. Once enough data has been successfully generated, the processing phase starts. The ML algorithms require an initial processing step as they cannot work with raw data. In the first step, any categorical variable present in the dataset is handled. Precisely, variables such as employee training level and external cybersecurity advisor are mapped into numerical values using the one-hot technique, which are easier for the ML process.

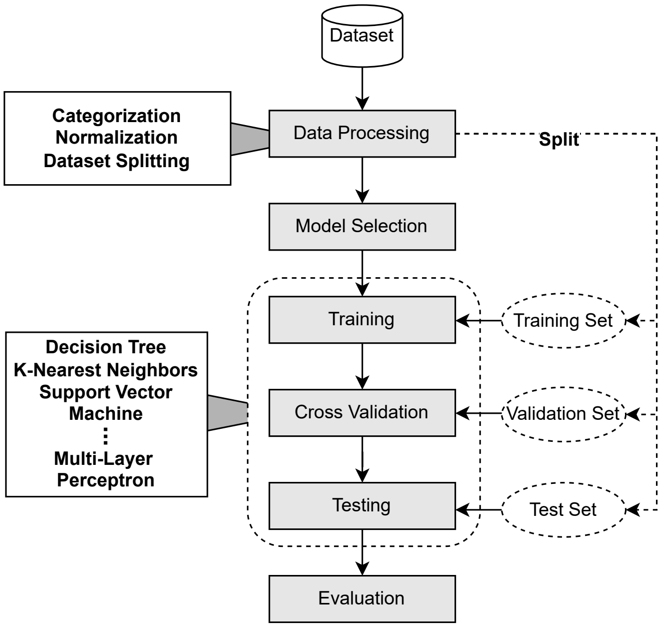


Fig. 2: ML Workflow Implemented by *SecRiskAI*

A further normalization may be necessary, depending on the selected ML algorithm. Normalization is the process of scaling data into a range of [*0*, *1*]. Some ML algorithms are susceptible to features with varying degrees of magnitude, range, and units. The dataset generated for *SecRiskAI* includes different features, such as revenue and number of employees with different ranges and training sensitive models on unscaled data may lead to lower performance and accuracy. Therefore, a normalization technique known as Min-Max scaling is used as defined by Equation 3. The Min-Max normalization technique is applied to the entire dataset but only to features (*i.e.*, every column except the risk), which contains the three output classes based on which future predictions will be made.

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{3}$$

The processing phase then involves splitting the dataset into a training, validation, and test set. After successful training and validation, the final model is subjected to extensive testing. During this phase, the final model is usually evaluated using previously unseen data (*i.e.*, test set generated using the implemented synthetic data generator), also called holdout set. The size of each dataset for training and test range from 5,000 to 50,000 (*cf.* Section IV).

B. Multi-Class Classification Algorithms

In ML, Multi-Class Classification (MCC) algorithms aim to solve problems of classifying instances into one of three or more output classes. Popular MCC algorithms are chosen for qualitative cybersecurity risk assessments in the model selection phase. A literature review was conducted on risk assessment to identify the most explored ML-based algorithms and their application scenarios. Based on that, four main algorithms were selected for *SecRiskAI* due to their favorable characteristics, application scenarios, and learning process. Thus, it allows to design and develop ML models that, based on contextual information, can make accurate qualitative risk assessment predictions and further monitor the organization's infrastructure by providing continuous assessment based on input data.

1) Decision Tree: Decision Tree (DT) is a Supervised Learning (SL) algorithm for the classification used in the proposed solution. This technique essentially looks at the feature values of the input dataset and categorizes them according to a specific parameter, also known as information gain.

The goal is to find, in a given dataset *D*, the feature having the highest information gain, which will in turn serve as a decision node of the tree. Next, the algorithm splits the dataset on the identified decision node and performs the search on the sub-datasets. A tree structure is then constructed, with each node representing a feature column and the leaves indicating the output class.

Besides being an easy-to-use and straightforward classification technique, this algorithm can be trained on historical data without requiring extensive data pre-processing. Compared to other classification algorithms used in *SecRiskAI*, the DT requires less effort for data preparation, and the normalization step is not required. Thus, the resulting model is easy to understand for both technical and non-technical stakeholders. In order to make a prediction using the DT, a new sample *i* would traverse the tree based on each feature value, and the resulting leaf value would be the output class.

2) K-Nearest Neighbors: K-Nearest Neighbors (KNN) is usually referred to as instance-based classifier as the main idea behind this technique is to memorize the input dataset to make future predictions. KNN requires three input parameters: a dataset *D* containing the historical information is given, a chosen number of neighbors *k* and *x*, a sample that is to be classified. The algorithm then proceeds on computing the distance between *x* and every record contained in *D*. Next, the computed distances are sorted in ascending order and *k* closest samples, also known as *neighbors*, to *x* are selected. Finally, the predicted class of *x* ($Class_x$) is based on the similarity with the neighbors, meaning that *x* is labeled following a majority voting of classes among the neighbors.

In essence, KNN calculates the probability of a sample *x* belonging to a specific class, based on neighbors observations. Compared to the DT, KNN requires more data pre-processing. On the other hand, the training phase is definitely faster and new training data can be seamlessly added without the need of reconstructing the model. If one supposes that the

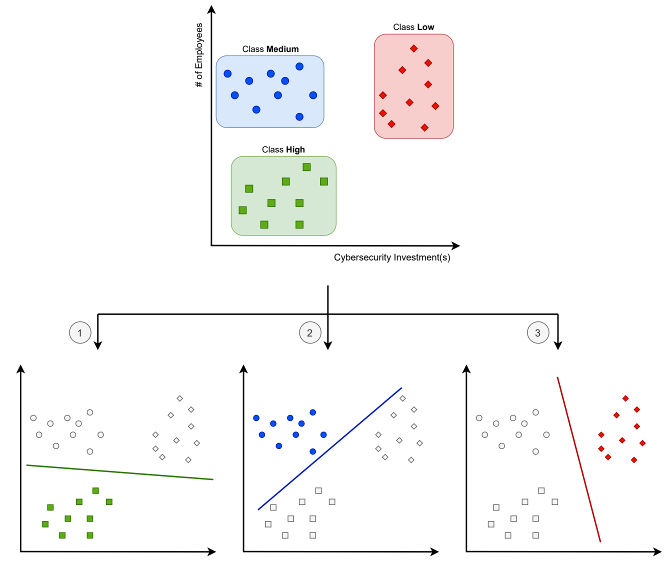


Fig. 3: SVM Visualization

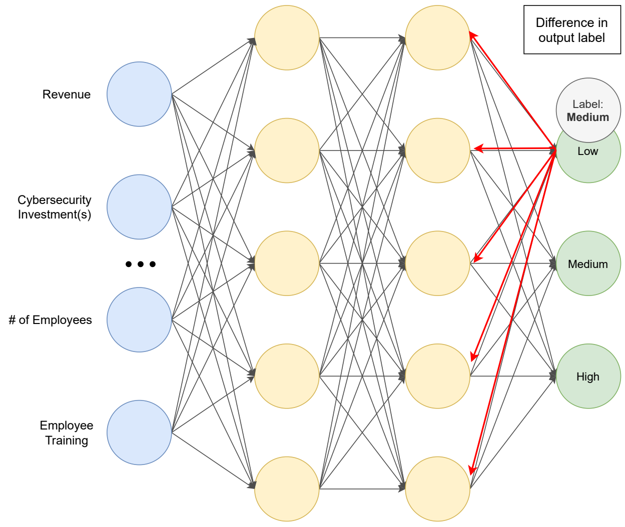


Fig. 4: MLP Visualization

representation of the KNN classification with *k* equal to seven and *x* being a new sample to classify. In this example, only two dimensions are taken into account (*i.e.*, cybersecurity investment(s) and number of employees). Once the *k* closest neighbors to *x* are identified, the predicted class of *x* is Low if the majority of the neighbors belong to the Low class.

3) Support Vector Machine: The Support Vector Machine (SVM) is the third SL classification algorithm implemented in *SecRiskAI*. In contrast with DT and KNN, SVM uses a line or hyperplane to separate input data into classes. Moreover, SVM is known to be computationally less expensive than KNN but does not support MCC natively. To achieve that, a *One-vs-Rest* strategy is followed. First, the multi-class dataset is broken down into multiple binary classification problems as highlighted in Figure 3. In this case, the following classification problems are identified:

- High vs {Low, Medium} (Figure 3 - Step 1)
- Medium vs {Low, High} (Figure 3 - Step 2)
- Low vs {Medium, High} (Figure 3 - Step 3)

Next, a binary classifier is trained on each binary classification problem and is able to predict a class probability (P_{class}), *i.e.*, the probability of an object belonging to a specific class. After the training phase, the binary classifiers return the probability of a sample being labeled as Low (P_{Low}), Medium (P_{Medium}) and High (P_{High}). Finally, the model that is able to predict the class of an unclassified sample *x* with the highest confidence is chosen and is represented in Equation 4:

$$Class_x = argmax(P_{Low}, P_{Medium}, P_{High}) \tag{4}$$

When dealing with larger datasets and n output classes, SVM would require the creation of n binary classifiers for each class, resulting to high computational costs. SVM does also suffer from performance issues when confronted with overlapping classes, *i.e.*, data points being not well separated. In contrast, SVM is a very flexible algorithm and allows the specification of a kernel function that can be linear (*cf.* Figure 3) but can also be of different types, such as nonlinear, polynomial, radial basis function, and sigmoid to solve many non-linear problems.

4) Multi-Layer Perceptron using Backpropagation: Multi-Layer Perceptron (MLP) is also explored in *SecRiskAI*. More specifically, MLP is a class of feedforward ANN. Figure 4 gives a visual representation of the MLP model implemented for *SecRiskAI*. Each node in the input layer corresponds to a specific feature of the generated dataset. Moreover, the MLP model has a total number of two hidden layers having five neurons each. Choosing the best parameters for an ANN is a very challenging task, as there are no clear rules and it really depends on the complexity of the underlying problem. The decision was based on the general guidelines available on the literature as well as extensive exploratory research and testing. Meanwhile, the output layer was defined based on the output classes of the model (*i.e.*, Low, Medium, and High). Therefore, it consists of three neurons representing each possible classification state.

During the training phase, the MLP uses a technique called *backpropagation*. An ANN propagates the input data forward through the neurons towards the output layer, where the prediction occurs. The backpropagation algorithm refers to the process of propagating the information about the prediction error backward from the output layer throughout the entire network, with the goal of adjusting the weights and improve accuracy. Figure 4 also depicts an example of a backpropagation mechanism initiated as soon as the original label (Medium) and predicted class (Low) differ. The computed error/loss is calculated and lastly is used to adjust the weights in the hidden layers.

Once the dataset is generated and the required ML algorithms are chosen, the training phase is initiated. Examples of the training datasets and overall process are available at [3]. First, the dataset is split following the 80-20 train-test strategy. Next, the process of choosing a set of optimal hyperparameters, also called hyperparameter optimization, takes place. The main idea is to use grid search to test every combination from a pre-defined list of parameters values (*cf.* Table I) required by the ML algorithm to build the model. Subsequently, the performance of each model is evaluated with the help of a 5-fold Cross Validation (CV) strategy. The model with the highest accuracy is selected and tested with unseen data (*i.e.*, the test set). Lastly, the entire process is applied to each ML algorithm discussed in the previous sections.

IV. Evaluation

SecRiskAI approach was evaluated considering two dimensions: *(a)* a quantitative evaluation to analyze the performance and effectiveness of the proposed algorithms and the features selection, and *(b)* a qualitative evaluation based on a case study to show evidence of the feasibility and usability of the *SecRiskAI* Web-based interface.

A. Performance Evaluation

The performance evaluation was conducted on a machine using the Apple M1 System on a Chip (SoC) and 16 GB of RAM. For a quantitative evaluation and comparison of the various ML models, the performance metrics of accuracy, precision, recall, and F1-Score were observed. The generation of these metrics is also a very important step in every ML workflow for understanding the behaviour and performance of the implemented models.

Confusion matrix is a widely adopted technique to evaluate the correctness and accuracy of classification models. In practice, confusion matrices can be used for both binary and multi-class classification problems and provide a way to assess and compare the performance of classification models. To compute the confusion matrix, a dataset containing 50,000 entries was generated and a 80-20 train-test split strategy was followed. Examples of the training datasets for overall cyberattacks and DDoS attacks are available in the *Training Dataset* folder provided at [3].

After the training and CV phase, each ML model was tested using the remaining test set. The purpose of this phase is to test the model on previously unseen data, *i.e.*, data not used during the training phase. For each entry in the test set, the ML model is used to predict the corresponding class. Finally, the predicted labels are compared with the actual class, also called true label, and as result a confusion matrix is built. Figure 5 shows the confusion matrices generated for each ML algorithm implemented in *SecRiskAI.*

A confusion matrix is essentially a summary of prediction results where each cell corresponds to the number of correct and incorrect predictions, broken down by predicted/true label combination. A best-performing classifier would result in a confusion matrix where only the diagonal is filled with values, meaning that every predicted class corresponds also to the actual label. In that case, the model would have achieved an accuracy of 100%. In other words, the accuracy of a model is calculated as the number of correctly predicted classes divided by the incorrect predictions. As shown in Figure 5, for the *SecRiskAI* prototype, every model was able to achieve more than 90% of accuracy. However, as can be observed in the cells near the diagonal, DT and KNN scored slightly worst than SVM and MLP.

Table II shows the computed performance metrics, based on the generated dataset with 50,000 entries. Each model was trained and tuned to maximise accuracy, reduce overfitting, and provide better results. SVM and MLP achieved similar accuracy scores, although in terms of computation time the difference is substantial. As for the training phase, SVM requires approximately half of the time compared to the MLP model. The training time has also an impact on the grid search computation time, a hyperparameter technique, which for the MLP model exceeds *200* seconds, since every tuned model undergoes a 5-fold CV. On the other hand, DT and KNN have the fastest training time, while KNN achieved the fastest grid search computation time of around *40* seconds.

TABLE II: Performance Metrics

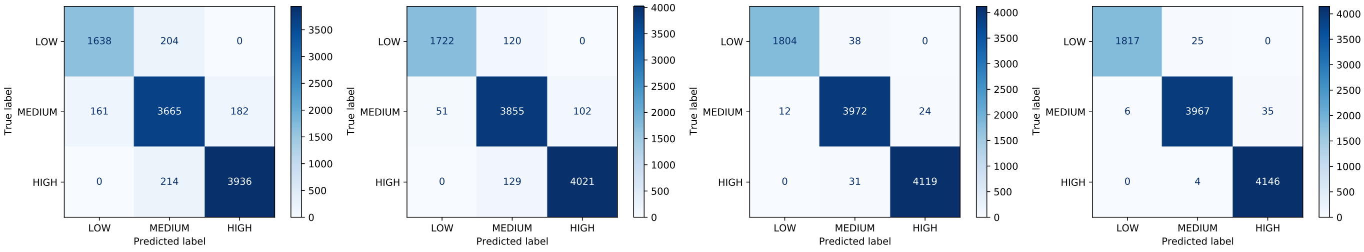
ML Model	**Accuracy**	**Training Time (s)**	**Grid Search Computation (s)**
DT	92.64%	0.18	146.77
SVM	99.03%	5.83	149.15
KNN	95.82%	0.08	40.06
MLP	98.86%	10.53	210.55

Based on the confusion matrices presented in Figure 5, the important metrics of precision, recall, and F1-score were derived as well. The precision metric is used to express the proportion of units labeled by a model that actually belong to that class. As shown in Figure 5 (a), DT was able to predict a *Low* risk for 1638 profiles out of all predicted profiles (1638 + 161 + 0), resulting in a precision of (1638 / 1799) ≈ 91%.

Additionally, the recall metric quantifies a model's predictive accuracy for a particular class, *i.e.*, it represents the ability of a model to find all entries in a dataset that belong to a particular output class. As presented in Figure 5 (a), out of 1842 (1638 + 204 + 0) profiles with *Low* as a true label, DT was only able to classify 1638 correctly, resulting in a ≈ 89% (1842 / 1638) recall.

The last performance metric considered in this evaluation is the F1-score, which ranges between *0* and *1*. This metric aggregates both precision and recall by computing the harmonic mean and is used to compare ML models to determine which one produces the best results. Similar to precision and recall, F1-Score is computed for each output class. Table III shows an overview of the derived performance metrics calculated for each ML model.

Additionally, the computed performance metrics summarized in Table III reveal that MLP, despite having a marginally lower accuracy than SVM, was able to achieve an F1-Score of *1.0* for the *High* output class. MLP also marginally outperformed SVM in both precision and recall scores. The small performance gain comes at cost of training time which,



(a) DT

(b) KNN

(c) SVM

(d) MLP

Fig. 5: Confusion Matrices for the *SecRiskAI*'s ML Model

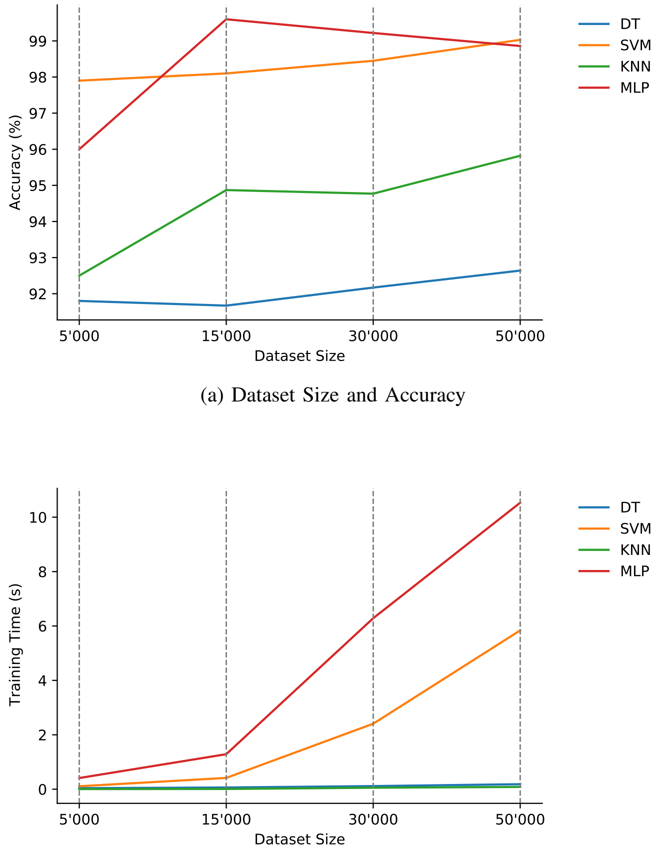
TABLE III: Computed Precision, Recall, and F1-Score for Each ML Model

ML Model	Class	Precision	Recall	F1-Score
DT	Low	0.91	0.89	0.90
	Medium	0.90	0.91	0.91
	High	0.96	0.95	0 95
SVM	Low	0.99	0.98	0.99
	Medium	0.98	0.99	0.99
	High	0.99	0.99	0.99
KNN	Low	0.97	0.93	0.95
	Medium	0.94	0.96	0.95
	High	0.98	0.97	0.97
MLP	Low	1.00	0.99	0.99
	Medium	0.99	0.99	0.99
	High	0.99	1.00	1.00

according to Table II, is generally higher compared to SVM classifiers. This is mainly due to the higher complexity of the MLP algorithm. The other two ML models used by *SecRiskAI* (DT and KNN) were also able to achieve still fairly high precision, recall, and F1-Score. However, similar to the accuracy scores presented in Table II, the metrics presented in Table III confirmed once again that DT provided the worst performance, despite having faster training times.

Finally, the impact of the size of the synthetic datasets on both accuracy and training time was investigated. First, datasets of different sizes were generated. Each ML model was equally trained and tuned on every generated dataset using grid search followed by a 5-fold CV technique. The results are shown in Figure 6. For small to medium size datasets (*i.e.*, between 5,000 and 15,000), the MLP model is able to outperform every other model with an accuracy of almost 100% (Figure 6 (a)). Moreover, the impact on the training time is also relatively low, with MLP requiring approximately *2* seconds. However, the outcome is different once the size of the dataset increases. On the one hand, the training time for both SVM and MLP increases drastically, which for MLP leads to a ≈ 388% increase with double the dataset size. On the other hand, as highlighted by Figure 6 (b), the accuracy of the MLP model suffers a slight decrease while having the highest accuracy score among the other ML models.

Once the generated dataset size reaches over 50,000 entries, MLP starts to perform worst than SVM while requiring twice as much time to be trained. While this may not seem to be a significant difference in terms of seconds, with datasets exceeding millions of entries, the gap may become even more substantial, leading to extremely slow model training and poor scalability. Furthermore, from Figure 6 (a), it can be observed that, with larger dataset sizes, SVM has a minimal but constant increase in accuracy. Similarly, KNN and DT also experienced an accuracy gain while maintaining a low training time. Therefore, based on the Figure 6, it is possible to conclude that MLPs really exceed with medium-sized datasets and SVMs should be taken into consideration when dealing with large datasets.



(a) Dataset Size and Accuracy

(b) Dataset Size and Training Time

Fig. 6: Dataset Size Evaluation

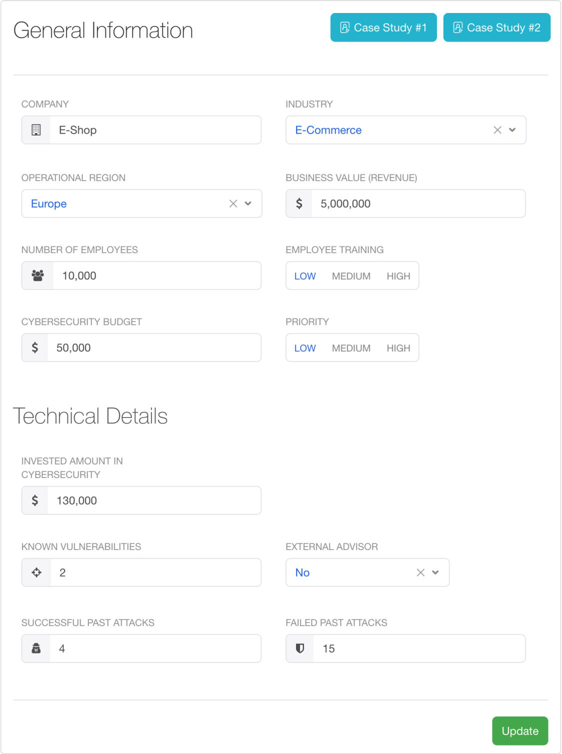
B. Case Study: Prediction of DDoS Attacks

This case study investigates and analyzes the ability of *SecRiskAI* to assess DDoS risks. For this purpose, an ML model was trained on a different dataset, adjusted to generate a different set of attributes that directly impact the likelihood of a company being targeted by DDoS attacks. After an exploratory research to understand behavior of DDoS attacks [7], attributes indicating the industry sector and the operative region as contextual information were defined. It happens because the number of DDoS attacks tend to be higher on telecommunications, financial, and sales service [9]. Also, Europe, the Middle East, and Africa accounted for more than half of the world's DDoS attacks [20]. Other attributes, such as the number of employees and employee training, were discarded and not included in the generation step because the impact of those on the DDoS risk has not been assessed by other studies. Consequently, based on the qualitative evaluation previously presented, a dataset of 30,000 entries was generated and MLP was chosen as the model for predicting DDoS risk. As most of these attacks reported contain multiple vectors (*i.e.*, not relying on just one type of DDoS attack), this case study does not consider one specific type of DDoS (*e.g.*, SYN, ICMP, or UDP flood), but rather the overall risk of having services affected by a DDoS attack.

As for the contextual information, it was assumed that the company interested in assessing DDoS risks was operating in the E-Commerce sector, buying and selling various types of goods over the Internet and mainly focusing on the European market. Moreover, the number of employees is around 10,000 and their training level, also understood as *awareness level*, for cybersecurity-related topics was classified as *low*. As shown in Figure 7, other general information includes a business value of around US$ 5M and a cybersecurity budget of just US$ 50,000. In this particular case study, the cybersecurity budget is intended to be used in either protection services of proactive/reactive nature or other investments aiming to increase DDoS resiliency.

In addition to the general information, the ML classifier requires some technical details as well. More specifically, the company has to provide the amount of US$ already invested in cybersecurity as well as any known vulnerabilities, which may derive from third-party security tools and penetration tests. Additionally, the number of failed/successful past DDoS attacks must also be reported, and the presence of an external cybersecurity advisor must be indicated as well.

The profile is updated and submitted to the *SecRiskAI*'s backend for further processing. The user can navigate the dashboard through the sidebar, where the contextual information is presented, and the risk prediction is automatically triggered. The middleware processes the company's profile, while the ML classifier delivers the actual prediction. The prediction response is rendered by the frontend and integrated into the dashboard. The overall DDoS risk prediction for the given profile here is *Medium*.



General Information

COMPANY: E-Shop

INDUSTRY: E-Commerce

OPERATIONAL REGION: Europe

BUSINESS VALUE (REVENUE): $ 5,000,000

NUMBER OF EMPLOYEES: 10,000

EMPLOYEE TRAINING: [✓ LOW] MEDIUM HIGH

CYBERSECURITY BUDGET: $ 50,000

PRIORITY: [✓ LOW] MEDIUM HIGH

Technical Details

INVESTED AMOUNT IN CYBERSECURITY: $ 130,000

KNOWN VULNERABILITIES: 2

EXTERNAL ADVISOR: No

SUCCESSFUL PAST ATTACKS: 4

FAILED PAST ATTACKS: 15

Fig. 7: Contextual Information for the DDoS Case Study defined in the *SecRiskAI* Web-based Interface

C. Discussion and Limitations

An important limitation of *SecRiskAI* is the current lack of real-world datasets for training the used ML algorithms. To partially circumvent this issue and show the effectiveness of the prototype, a synthetic dataset generation approach was followed. However, although synthetic data mimics various properties and aspects of real-world data, it is usually very challenging to generate high-quality data for complex problems. If the generated dataset does not match the behavior and properties of the real-world dataset, this will negatively impact on the performance of the trained ML models. Also, the current implementation of *SecRiskAI* supports assessing the risks of DDoS and Phishing attacks only. Still, the system prototype is extensible so that new ML models, trained explicitly considering other cyberattacks, can be integrated into the current solution and exposed through the same API. For that, the same ML workflow defined for *SecRiskAI* can be followed.

The quantitative evaluation of these four ML algorithms demonstrated that SVMs achieve a higher accuracy for larger datasets, while maintaining a lower training time when compared to MLP. Nonetheless, all ML algorithms performed well and achieved more than 90% of accuracy in most cases with the correct training dataset. The generated confusion matrices also confirmed that these ML algorithms classify most samples correctly. Other important metrics (precision, recall, and F1-Score) also provided valuable insights about the performance of the ML algorithm for every output class.

V. Conclusions and Future Work

SecRiskAI provides a clear ML workflow to design and implement an ML-based tool for supporting the process of cybersecurity risk assessment in companies. First, the overall ML risk assessment workflow was designed, encapsulating essential steps of risk predictions, such as data gathering, data processing, ML model selection, and performance evaluation. Specifically, this work investigated the suitability of four ML algorithms (*i.e*., DT, SVM, KNN, and MLP) of predicting and assessing the likelihood of cyberattacks. *SecRiskAI* is the first of its kind to propose an ML-based approach for the correlation of business attributes for the risk assessment of economic impact in businesses. However, further investigations are still required to determine which are the best information to consider for real-world companies.

SecRiskAI's prototype [3] implements two specific ML models to assess the risk of DDoS and Phishing attacks. In order to show the feasibility of the proposed ML workflow, *SecRiskAI* implements two ML models to predict the risks of being targeted by either DDoS or Phishing attacks. In addition, *SecRiskAI* prototype allows the integration with recommender systems [5], [6] to provide a list of recommended protection services based on the business profile and also influenced by the calculated risk.

Future work includes investigating the relevance and weight of each data attributes for the learning process in order to refine cyberattack-specific ML models. These models can be fully specialized to assess the risk of specific cyberattacks, allowing for a more comprehensive cybersecurity risk assessment phase. Explainable AI techniques can also be considered to better understand to which extend the selected data attributes influence the risk assessment results. Lastly, further experimental tests are needed to evaluate the behavior and performance of the ML models currently implemented in *SecRiskAI* on real-world scenarios and datasets.

Acknowledgements

This paper was supported partially by *(a)* the University of Zürich UZH, Switzerland and *(b)* the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 830927, the CONCORDIA project.

References

[1] P. Avesani, A. Perini, A. Siena, and A. Susi, "Goals at Risk? Machine learning at Support of Early Assessment," in *IEEE 23rd International Requirements Engineering Conference (RE 2015)*, Ottawa, Canada, 2015, pp. 252–255.

[2] Y. Castro and Y. J. Kim, "Data Mining on Road Safety: Factor Assessment on Vehicle Accidents using Classification Models," *International Journal of Crashworthiness*, vol. 21, no. 2, pp. 104–111, 2016.

[3] E. Sula, M. Franco, "SecRiskAI - Source Code," 2021, https://gitlab.ifi.uzh.ch/franco/ml-risk-smes.

[4] European Union Agency for Cybersecurity (ENISA), "Threat Landscape," October 2020, https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends.

[5] M. Franco, B. Rodrigues, and B. Stiller, "MENTOR: The Design and Evaluation of a Protection Services Recommender System," in *15th International Conference on Network and Service Management (CNSM 2019)*. Halifax, Canada: IEEE, October 2019, pp. 1–7.

[6] M. Franco, E. Sula, B. Rodrigues, E. Scheid, and B. Stiller, "ProtectDDoS: A Platform for Trustworthy Offering and Recommendation of Protections," in *Economics of Grids, Clouds, Systems, and Services*. Izola, Slovenia: Springer International, 2020.

[7] M. Franco, J. Von der Assen, L. Boillat, C. Killer, B. Rodrigues, E. J. Scheid, L. Granville, and B. Stiller, "SecGrid: a Visual System for the Analysis and ML-based Classification of Cyberattack Traffic," in *IEEE 46th Conference on Local Computer Networks (LCN)*, 2021, pp. 140–147.

[8] M. Kalinin, V. Krundyshev, and P. Zegzhda, "Cybersecurity Risk Assessment in Smart City Infrastructures," *Machines*, vol. 9, no. 4, 2021.

[9] Kaspersky Lab, "Denial of Service: How Business Evaluate The Threat of DDoS Attacks," 2021, https://media.kasperskycontenthub.com/wp-content/uploads/sites/45/2018/03/08234158/IT_Risks_Survey_Report_Threat_of_DDoS_Attacks.pdf.

[10] B. A. Khalaf, S. A. Mostafa, A. Mustapha, M. A. Mohammed, and W. M. Abduallah, "Comprehensive Review of Artificial Intelligence and Statistical Approaches in Distributed Denial of Service Attack and Defense Methods," *IEEE Access*, vol. 7, pp. 51 691–51 713, 2019.

[11] V. S. Kumar and V. L. Narasimhan, "Using Deep Learning For Assessing Cybersecurity Economic Risks In Virtual Power Plants," in *7th International Conference on Electrical Energy Systems (ICEES)*, Chennai, India, 2021, pp. 530–537.

[12] V. S. Mai, R. J. La, and A. Battou, "Optimal Cybersecurity Investments in Large Networks Using SIS Model: Algorithm Design," *IEEE/ACM Transactions on Networking*, vol. 29, no. 6, pp. 2453–2466, 2021.

[13] S. Morgan, "Cybercrime to Cost The World $10.5 Trillion Annually By 2025," 2020, https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021.

[14] S. M. Pappalardo, M. Niemiec, M. Bozhilova, N. Stoianov, A. Dziech, and B. Stiller, "Multi-Sector Assessment Framework - A New Approach to Analyse Cybersecurity Challenges and Opportunities," in *Multimedia Communications, Services, and Security*. Krakow, Poland: Springer, Lecture Notes in Computer Science (LNCS), 2020, pp. 1–15.

[15] C. M. Rocco S. and E. Zio, "A Support Vector Machine Integrated System for the Classification of Operation Anomalies in Nuclear Components and Systems," *Reliability Engineering & System Safety*, vol. 92, no. 5, pp. 593–600, 2007, recent Advances in Theory & Applications of Stochastic Point Process Models in Reliability Engineering.

[16] B. Rodrigues, M. Franco, G. Paranghi, and B. Stiller, "SEConomy: A Framework for the Economic Assessment of Cybersecurity," in *16th International Conference on the Economics of Grids, Clouds, Systems, and Services (GECON 2019)*. Leeds, UK: Springer, September 2019, pp. 1–9.

[17] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," *IEEE Access*, vol. 8, pp. 222 310–222 354, 2020.

[18] P. M. S. Sánchez, J. M. J. Valero, A. H. Celdrán, G. Bovet, M. G. Pérez, and G. M. Pérez, "A Survey on Device Behavior Fingerprinting: Data Sources, Techniques, Application Scenarios, and Datasets," *IEEE Communications Surveys Tutorials*, vol. 23, no. 2, pp. 1048–1077, 2021.

[19] M. Taamneh, S. Alkheder, and S. Taamneh, "Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates," *Journal of Transportation Safety & Security*, vol. 9, no. 2, pp. 146–166, 2017.

[20] W. Ashford, "Europe in the Firing Line of Evolving DDoS Attacks," 2018, https://www.computerweekly.com/news/252434746/Europe-in-the-firing-line-of-evolving-DDoS-attacks.

[21] Q. Zhang, C. Zhou, Y.-C. Tian, N. Xiong, Y. Qin, and B. Hu, "A Fuzzy Probability Bayesian Network Approach for Dynamic Cybersecurity Risk Assessment in Industrial Control Systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2497–2506, 2018.

[22] Q. Zhang, C. Zhou, N. Xiong, Y. Qin, X. Li, and S. Huang, "Multimodel-Based Incident Prediction and Risk Assessment in Dynamic Cybersecurity Protection for Industrial Control Systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 10, pp. 1429–1444, 2016.

All links provided above were last accessed on May, 2022.