University of Zurich[UZH]

MASTER PROJECT – Communication Systems Group, Prof. Dr. Burkhard Stiller

CyberAlert: An ML-based Cybersecurity Risk Assessment Tool

Chenfei Ma, Euxane Vaz Pinto, Neng Xu
Zürich, Switzerland
Student ID: 21-740-816, 18-211-391, 21-740-985

Supervisor: Dr. Muriel Franco, Dr. Alberto Huertas
Date of Submission: January 16, 2023

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

Master Thesis
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Abstract

The exponential increase of cyberattacks has been getting more and more worrying over the years. The shift to a more online-oriented world during and after the Covid-19 pandemic puts to light the impact of cyber-issues on the good functioning of enterprises, users and society as a whole. The confidentiality surrounding the cybersecurity field prevents researchers from working with real data coming from companies. This project builds on top of other previous works in an attempt to bring more awareness to companies on their own vulnerability to cyberattacks. To start, a synthetic dataset is carefully generated and then used to train different machine learning models, in order to predict the cybersecurity risk entailed by companies. In the attempt to approximate real-life data, different types of noise are injected into the data and in the label generation. The models are then evaluated and the reasons behind their performance discussed. Besides, to give users access to the models, the tool CyberAlert was developed. This tool can serve as an intuitive and beginner-friendly platform to get insight into the risk of a company, or be personally configured by more advanced users.

Acknowledgments

We would like to sincerely thank our supervisor, Dr. Muriel Franco for his constant support, help, guidance and positive energy throughout the development of this master project.

We are also thankful to our co-supervisor, Dr. Alberto Huertas, as well as to Prof. Dr. Burkhard Stiller and to the CS Group members for their valuable feedback, and for the opportunity of working on this topic.

Contents

Chapter 1

Introduction

The field of Cybersecurity is constantly evolving, getting always more detailed, technical and difficult to handle due to the amount of information and systems involved. Different challenges and solutions emerge regularly [11, 12] while cyberattacks get more and more refined and cybercriminals always end up finding ways to exploit individual and business vulnerabilities. Patching these vulnerabilities when they are discovered is the most adequate approach, and being able to handle or even anticipate a cyberattack or risk is crucial. According to the Allianz Risk Barometer of 2021[7], over 1 trillion of global losses have been caused by cybercrime. Therefore, it would be more beneficial for businesses to be able to proactively prevent attacks before they occur and organize their defense accordingly.

Especially during and after the COVID-19 pandemic, the challenges that represent cybersecurity and privacy have become even more outstanding [59]. Remote working went from an odd rule, necessary in regards to the safety of the population to a practice that now seems normal to most, and will likely be still widely used by most companies [2]. This new way of approaching the work-life impacts strongly not only the lives of employees but the technical organization of the enterprise as well and brings its share of new challenges concerning security. Since the pandemic the number of cybersecurity complaints the Federal Bureau of Investigation (FBI) of the United States received went up by 300%, and Google noticed an increasing amount of phishing emails [3]. The outbreak of COVID-19 might have been a catalyst for cybercrime these past few years, but this does not imply that cybersecurity issues stem from the pandemic. Before the health crisis, the number of attacks was already increasing each year, and the cybercriminals' methods adapted to target the current vulnerabilities [9, 10].

1.1 Motivation

A critical characteristic of cybersecurity is the lack of transparency of businesses that have been victims of cyberattacks. Whether it is motivated by the fear of losing the trust established with their customers, or loss of market shares and reputation among other

enterprises, businesses do not like to report cyberattacks. A study showed that managers usually are open to admitting smaller breaches, but choose not to disclose attacks that strongly harmed their business unless their investors might be already aware of them [6]. It is understandable that getting damages from a cyberattack could tarnish the reputation of an enterprise, and one can be easily tempted to minimize the damages to the public's eyes in an attempt to keep their business' credibility. On top of that, releasing public information on how a vulnerability was exploited and its impact is clearly not aligned with companies' security guidelines.

However, this behavior concerning sharing cyberattacks experience is eventually detrimental to the enterprises. The low amount of information surrounding this matter makes it difficult for businesses to estimate if they are themselves vulnerable to cyberattacks since they cannot be fully aware of the cybercrimes currently being used, how the vulnerabilities are exploited nor make comparisons with their own situations. As enterprises can only have access to a really scarce amount of outside information to refine their defense against cyberattacks, they are left with being only able to rely on themselves - or external consulting to organize and keep their cybersecurity measures up to date. If the bigger enterprises can afford to have a lot of resources dedicated towards cybersecurity and risk assessments, it is not always the case for Small and Medium-sized Enterprises (SMEs) with less financial and human resources. Although their general awareness of the importance of cyber threats, the majority of SMEs are reluctant to make cuts in the budget to allocate more capital to reinforce their investment in cyber security [8].

To get a better understanding of the underlying risks which businesses are facing, different threat modeling and risk assessment approaches are in use [30, 29]. These approaches are not only tied to the world of cybersecurity as they can help to assess the variety of hazards that enterprises can come across. Among others, Machine Learning (ML) models can be used to provide a prediction of possible cybersecurity risks and their underlying costs [13]. It is important to raise that the key aspect of any ML model is having access to data. Different techniques can be used to get around specific situations, such as only disposing of unlabeled data. However, it is a non-negotiable condition to have some amount of data as input to be able to create, train and evaluate ML models.

1.2 Description of Work

This Master Project focuses on applying ML techniques to address different risk assessment challenges, such as the lack of information, lack of cybersecurity experts, and limited budget to perform complex tasks. For that, the SecRiskAI approach [4, 5] is used as a basis and initial step for this Master Project. Thus, this work provides a simplified approach to understanding possible risks a business could face due to cyberattacks. The contributions of this Master Project include *(i)* an investigation and refinement of SecRiskAI models, *(ii)* analysis of additional parameters of cybersecurity to propose new ML models for cybersecurity risk assessment, including those parameters related to threats, systems, and the business cybersecurity, *(iii)* generation of new training datasets based on real-world data and publicly available reports, and finally *(iv)* the development of a web-based interface for the interaction and training with models for risk assessment. Also, the evaluation

of proposed models is evaluated in this Master Project, including performance metrics, explainability of the models, and application scenarios.

Due to the lack of publicly available data sets on cyberattacks, it can be understood that only a few studies have been trying to find ML models for risk assessment in the domain of cybersecurity. Often, they make the decision to generate themselves a data set to train the algorithm. This is in particular the case for SecRiskAI. Generating a realistic data set is not an easy task. It needs to be the closest possible to the real world for it to provide a good basis for the algorithms and models. Biased training data will lead to the algorithm learning the wrong pattern, which further leads to outputting erroneous results when using the model with real data. Therefore, one of the challenges this Master Project addresses is the definition and refinement of data sets for ML-based cybersecurity risk assessment.

1.3 Report Outline

Chapter 2 consists of three parts. The first part is the introduction of Cybersecurity and Cyberattacks. The second part introduces three ML models used in the project. Finally, the last part discusses and analyzes related work in different ML fields.

Chapter 3 provides the main approaches implemented in this project, including the synthetic dataset generation and the development of a user interface.

Chapter 4 presents the evaluation and analysis of different models.

Finally, Chapter 5 summarizes the Master Project and provides future work for further study.

Chapter 2

Background and Related work

The following section introduces the theoretical foundation for the understanding of this work. It includes basic concepts of cybersecurity, risk assessment and its frameworks, cyberattacks as well as ML methods, in particular Random Forest, Support Vector Machine and Artificial Neural Networks. Then, a literature review is provided to highlight different ML methods, techniques and implementation that were used in previous works for various sectors.

2.1 Cybersecurity and Risk Assessment

The term of *cybersecurity* has been in use since the end of the nineteen eighties [24]. After reviewing numerous definitions that have been given to this term and analyzed in what way they were lacking, Craigen, Diakun-Thibault and Purse defined cybersecurity in their article in Technology Innovation Management Review as " [...] the organization and collection of resources, processes, and structures used to protect cyberspace and cyberspace-enabled systems from occurrences that misalign de jure from de facto property rights." [23]. Cybersecurity is a broad term that covers a large spectrum of practices, with its central aspect being the protection of the information and data of an individual or organization in the digital world.

Risk assessment is a major step of cybersecurity risk management, which provides measurable information to be used during the cybersecurity decision-making process. At this step, threats and vulnerabilities are identified, and risk is then calculated, usually as a function of the likelihood that these harmful events would occur. Results of risk assessment are used in a further step of the risk management process. They help ensuring that an organization is prioritizing the right aspects and taking relevant countermeasures to respond to the identified risks [19].

Different frameworks offer different approaches for risk assessment, each focusing particularly on certain aspects, such as technical and economic [31, 33]. The Operationally Critical Threat, Asset and Vulnerability Evaluation (OCTAVE) approach consists in exploiting an enterprise's knowledge of its own security practices, and targets mainly organisational

risks instead of the technical ones [20]. The National Institute of Standards and Technology (NIST) Cybersecurity Framework [19] is on the other hand a more straightforward approach, less time-consuming to implement. It provides a more general guideline that can be used by most enterprises, including the ones less familiar with risk management [21, 22].

The notion of risk can be interpreted in different ways, depending on the kind of information one needs to retrieve. This Master Project covers risk as the potential that an enterprise would be the victim of a cyberattack. Therefore, the risk assessment will be measured as the likelihood that an attack would be successful - and not the enterprise's probability of being targeted. The magnitude of the damage generated by that successful attack is beyond the scope of this work.

2.2 Examples of Cyberattacks

This section provides an overview of two common types of prominent cyberattacks: Phishing and Distributed Denial-of-Service (DDoS). These cyberattacks bring some insights into the cybersecurity risk equation and the development of ML models in the following sections.

Phishing is one of the social engineering attacks where criminals are using scams to trick and steal any person's identity or credential or install malware on their computers. In the past few years, phishing became one of the top threats to internet security. According to a survey in 2011, more than 50% of the total attack incidents are reported as phishing attacks and over 62% of organizations were victims of phishing attacks [32]. Nowadays phishing attacks are even more sophisticated as they include not only sending spam by email but also every aspect of communication like Short Message Service (SMS), telecom and social networking. They are also being more specifically targeted than ever before by using a technique known as spear-phishing, which uses contextual information related to specific individuals to make the attack hard to notice. Therefore, the more information is made available to the attackers, the more dangerous an attack will be.

Since phishing is a social engineering attack and thus is almost impossible to prevent by technical means, one of the most important ways to protect the company from such an attack would be phishing awareness training. Generally, phishing training is about teaching employees how phishing attacks work, how to identify a phishing email and how to be alert to the potential exposure of personal information.

DDoS attacks are another major threat and probably the hardest to detect and mitigate in cybersecurity [34, 35]. It essentially limits the access to the service by overwhelming the target with a flood of internet traffic. The target will be exhausted of computing and communication resources within a short period of time. The effect of a DDoS attack could also be severe. The most common and obvious one is the downtime of the website. With overwhelming network trafficking, your website could be unavailable until you fix the networking and potentially cause a lot of financial loss. Moreover, DDoS could lead

to more server issues and make your site even more vulnerable since all the focus is on getting the website working again.

However, there are several approaches to defend against a DDoS attack such as applying traffic filtering, intelligent routing techniques and implementing the server level DDoS protection. Since DDoS is a relatively simple yet powerful cyberattack method, it was one of the types of cyberattacks that was kept in mind in the design of the ML models.

2.3 Machine Learning Methods

The main goal of this paper is to design and develop ML models that, based on actual contextual information, can make accurate risk assessment predictions for companies based on input data. In supervised learning, there are many different algorithms aiming to solve Multi-Class Classification (MCC) problem. The following subsection list three of these algorithms, Random Forest, Support Vector Machine and Neural Networks. Due to the different properties they have, a performance and efficiency comparison of these ML algorithms is provided.

2.3.1 Random Forest

A Random Forest (RF) is an ensemble of Decision Tree (DT). DT is a non-parametric supervised learning method used for classification. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The general idea is to choose the best feature at each depth of the tree. The information gain is introduced to measure the value of different features:

$$IG(S, A) = Entropy(S) - \sum(\frac{|S_v|}{S} * Entropy(S_v)) \tag{2.1}$$

where S is the data feature, S_v are the values of the feature S.

During the training phase, each decision tree is trained with a subset of the training data that is sampled with replacement. In addition, a random subset of features may be used at each node. During the testing phase, the final class decision is either the most voted class by all the decision trees in the ensemble or the average of the predicted real values of all the trees in the ensemble.

2.3.2 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model for classification problems. It is trained to pick a separating boundary that maximizes the margin between the data points of different classes. For linearly separable cases, the optimization problem is convex

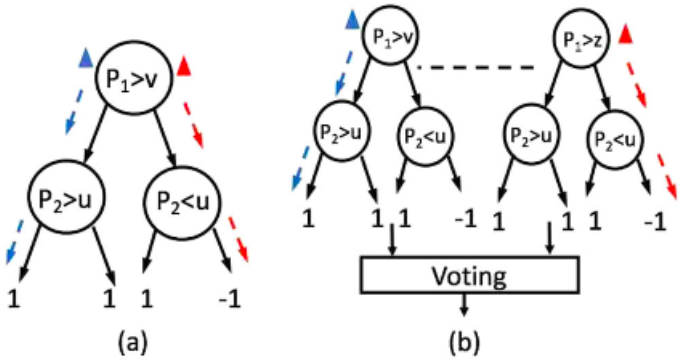


Figure 2.1: (a) Classification using DT; (b) Classification using RF [16]

quadratic; for non-separable cases, a slack variable is added to the problem so that a feasible solution can always be found.

A dual SVM formulation is also introduced to capture non-linear boundaries and applies the "kernel trick". There are different types of kernels used with SVMs such as the radial basis function (RBF) kernel, the Mercer kernel and the String Kernel. Kernel hyperparameters must be chosen carefully to improve the model accuracy. This basic SVM model has been extended to solve multi-classification problems and Support Vector Regression (SVR) for regression problems.

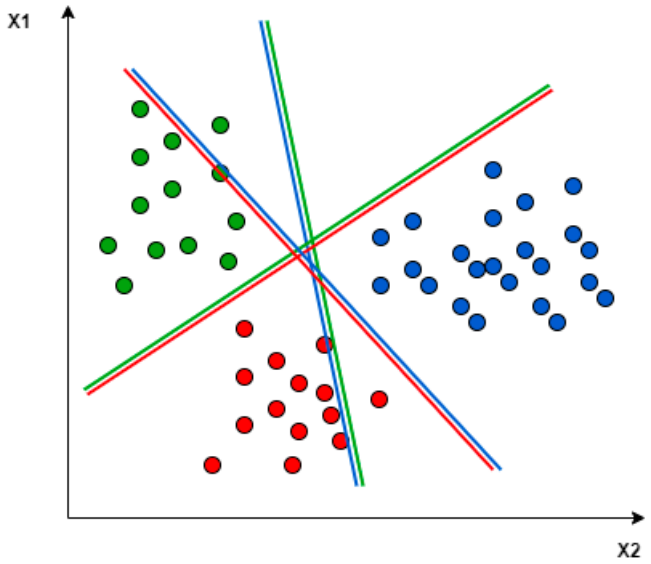


Figure 2.2: Multi-classification SVM in One-to-Rest approach [37]

2.3.3 Artificial Neural Network

Artificial Neural Networks (ANNs) mimic the brain by creating neurons that are interconnected. They can be used for classification, but also regression or clustering.

ANNs are organized in the following way. They start with an input layer which consists of the input data one wants the network to learn with. Then, there are one or multiple hidden layers which take the results of the previous layers as input and make it go through an activation function, e.g. a sigmoid function. As the activation function can be non-linear, this element introduces non-linearity to the network. It also helps to regulate by maintaining the outputs in an adequate range. The last layer is called the output layer and produces the final result of the ANN.

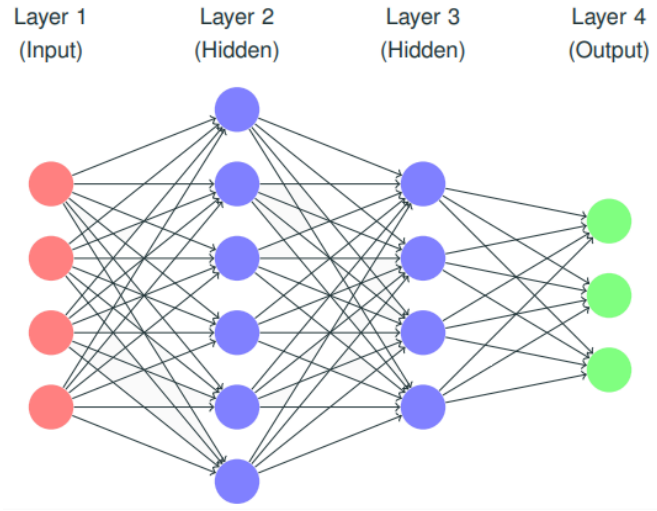


Figure 2.3: Layers of an Artificial Neural Networks [25]

There is a weight associated with all the connections - between each input neuron and each neuron of the hidden layer, between each neuron of the first layer and each neuron of the second layer, and so on.

During training, the results of the output layer of the ANN are compared to the expected results. When they are not close enough, the difference between the true result and the ANN output which is called the error is backpropagated to the weights. Adjusting these parameters is how the network learns the correct way to predict a result based on the given input [25, 26, 27].

2.4 Literature Review

This section provides a brief analysis of recent publications utilizing ML methods in engineering risk assessment as well as researches about adding ML in cybersecurity risk assessment.

2.4.1 Methods and Systems

As shown in Table 2.1, for the frequency of different methods, the ANN is the most often applied model, closely followed by SVMs which is also a classic classification model.

Jeevith Hegde et al. [13] provides a summary of the usage of machine learning methods in all risk assessment fields. Ma et al. [51] applies SVM and ANN to real-time highway traffic condition assessment, which both outperform a baseline California-type incident-detection algorithm. Also, other algorithms like DT and Bayesian Models are used to build a Safety-Route Planner in [53]. [56] develops an automated fall detection system by comparing six different ML techniques. The accuracy and specificity of different falls are all above 99%. In the Nuclear sector, [57] uses an ANN which consists of six inputs, two layers and one output to improve the condition-based maintenance (CBM) regime to detect faults in a nuclear power plant in transients. ANNs also perform best in [4], which compares four ML classification methods in predicting the risk of being attacked by different kinds of cyberattacks.

As for the types of implementations, in the automotive industry, all kinds of implementation such as case-study, real-world implementation, experimental tests, simulator tests, and review has been considered. In the field of cybersecurity, [16] provides an overview of opportunities for ML techniques in cybersecurity risk assessment. In [35] different ML techniques (K-Nearest Neighbour (KNN), RF, and ANN) are applied to analyze and classify DDoS attacks. Finally, [4] introduced SecRiskAI, an ML-based tool for cybersecurity risk prediction in companies.

Table 2.1: Industry-wise usage of ML Applications and Implementation for Risk Assessment

Industry	Work	Implementation	Used Algorithm
Automotive	VII system [51]	Real-world implementation	ANN, SVM
	Vehicle Accident [52]	Experimental tests	Bayesian, DT, ANN
	Route Planner [53]	Real-world implementation	ANN
	Topic Modeling [54]	Case study	RBF
Healthcare	Safety Perform [55]	Case study	Regression Model
	Wearable Sensors [56]	Experimental tests	KNN, ANN, SVM
Nuclear	Plant Simulator [57]	Experimental tests	ANN
	Dynamic PRA [60]	Real-world experiment	Clustering
Cybersecurity	Survey [58]	-	-
	SecGrid [35]	Experimental tests	RF, k-NN, ANN
	SecRiskAI [4]	Experimental tests	ANN, DT, SVM

2.4.2 Data Processing and Data Sources

Data processing is an important step before training the ML models. As for large datasets, [18] uses different kinds of feature selection techniques to reduce the initial attributes in aircraft accidents when facing a huge amount of knowledge and data collection. However, due to the lack of accessible data in the cybersecurity field, as shown in Table 2.2, data augmentation and data generation are needed to provide enough data to train the models proposed in this MAP. Il-Hwan Kim et al. [17] uses ANNs to augment the information on the vehicle states and the road surface condition, which is fed to SVM to detect the

driver's intention with high accuracy. But due to information confidentiality reasons, only two specific scenario cases of cyberattacks could be found. Although this is not enough for augmentation, they can be used for validation in the models evaluations when testing the performance of models.

Table 2.2: Datasets and Processing for Risk assessment in Different Topics

Topics	Datasets Source	Processing
Aircraft Accidents [18]	Aviation Companies	Reduction (CFS PCA)
Driver Assistance [17]	Vehicle states/Road Surface Condition	Augmentation
Road Risk Modeling [53]	Highway Safety Information System	Hybrid Neural Network
Detecting Falls [56]	Devices Experiments	Thresholding
Nuclear Plant [57]	Experiments through power transients	ANN
Cyber threat Analysis [1, 36]	Businesses Cybercrime	-
	Two real-world cases of cyberattacks	-
Cyber Risk Prediction [4]	Synthetic data	Data Generation

Another common way to generate data, especially in medical and physical fields, is to conduct experiments. [56] and [57] mainly get their data by real-time experiments: [56] develops an automated fall detection system with wearable motion sensor units fitted to the subjects' body recording six different positions and [57] builds a dynamic nuclear plant simulator. [4] studies the range of each attribute in cybersecurity risk assessment and uses a generalized equation to train the supervised learning models with synthetic data.

Chapter 3

Approach

The following chapter introduces the approach taken to train a model to predict the cybersecurity risk of enterprises. It first defines the attributes that have been deemed to be relevant to predict risk and gives some insight into the way they have been chosen. These attributes are further used to generate the dataset to train the models. Then, the method of generating labels for the generated dataset is detailed. Later, the training process of the ML models to predict labels on new data is presented. Finally, the tool for CyberAlert is introduced. Companies can use this intuitive user interface to enter their characteristics and get some insights into their own cybersecurity risk.

3.1 Attributes Definition

The first stage for the design of the risk prediction tool is to define which attributes will be taken into consideration. A choice has to be made between the different attributes that characterize enterprises. This choice is primarily produced in the function of the relevance of said attribute in relation to cybersecurity. Another element is whether an enterprise can accurately assess this information for itself. In the following table is the attribute that was selected to generate data that will be used to train the models. The range that they can take is shown next to them, followed by a short summary of the meaning of that attribute.

The attribute Access Control is divided into four levels. In the first case, the company does not use authentication, which further prevents implementing authorized access. The second one is the use of only one way of authentication, which enables authorized access. The last two entail Multi-factor Authentication, with either two or three authentication methods. Both also encompass authorized access. Not having any means of authentication means that anyone could easily access the enterprise's data and thus is a substantial vulnerability. Using one way of authentication - for example, a password - is already a step to improve security. In addition, having the users only authorized to access the files they are accredited for boosts the safety of the system. However, while only one way of authentication is better than none, it has been shown that MFA is necessary to significantly lower the risk of various types of cyberattacks [40].

Table 3.1: Attributes chosen to generate the synthetic dataset

Attribute	Measurement	Description
Access Control	0 = no authentication, no authorization 1 = basic authentication, authorization 2 = MFA (2), authorization 3 = MFA (3), authorization	Multi-factor authentication (MFA) has been shown to lower the risk of different types of cyberattacks. Authorization is already a good step to improve cybersecurity.
Cybersecurity Awareness	0 = low 1 = moderate 2 = high 3 = very high	It is important for the employees to understand the risks that they are subject to so that they can behave according to best practice.
Cybersecurity Investment	6 - 14%	Enterprises invest different amounts in cybersecurity, usually between 6 and 14 % of their IT budget. 13.7% is thought to be a good percentage to be allocated .
Data Storage	0 = outsourced cloud 1 = local cloud	Security measures undertaken by bigger cloud providers are likely to be more robust than local ones.
IT Support	0 = no professional IT support 1 = one or more IT experts 2 = IT security department	IT specialists help the company to take effective and quick measures to prevent cyberattacks.
Number of Employee	1 - 250	The number of employees is positively correlated with the number of social engineering vulnerabilities an enterprise is subject to. This is not an issue if they have undergone good training.
Revenue	1 - 50 millions €	A high revenue goes both ways: more money to invest into cybersecurity, but also more appealing to cybercriminals.
Attack Frequency (by industry)	0 = Education 1 = Government 2 = Business 3 = Health	Some industries have been found to be attacked more frequently than others.
Vulnerability Identification	1-10	There are a number of known vulnerabilities in an enterprise's system, that have been discovered but have not been patched yet.

As this model is constructed for SMEs, the maximum Number of Employees as well as the maximum Revenue have been delimited according to the threshold of what is considered an SME: until 250 employees, and up to a revenue of 50 million euros [44]. Both of these attributes are two-sided and can either raise or lower the risk, according to other parameters.

A higher Revenue means more money that can be invested to build a solid defense, but at the same time will spark cyberattackers' interest. This attribute is therefore closely linked to another one, Cybersecurity Investment: if sufficient resources are invested in protecting the company, then they will have the means to face this unwanted attention. While enterprises usually invest between 6% and 14% of their IT budget in cybersecurity, 13.7% is considered a good percentage to be allocated [39].

Similarly, a high Number of Employees can be mapped to a higher risk concerning social engineering vulnerabilities if the employees are lacking good cybersecurity behavior. It will however not be an issue if they have received good training to deal with such issues. Thus, this attribute is strongly linked with Cybersecurity Awareness. It measures to which degree the employees of a company are knowledgeable about the dangers of cyberattacks, and behave according to best practices. This attribute is important to consider as studies have shown that despite being highly concerned by the topic of cyber awareness, an important number of companies still rate the level of awareness of their own employees as low [12].

The chosen type of Data Storage also has repercussions on cybersecurity. There are different factors such as price, infrastructure or scalability that enterprises have to weigh in when deciding to store their data on external cloud storage or doing it locally. However, when looking at it from only the cybersecurity viewpoint, using the service of a bigger and more trustworthy company dedicated to cloud providing is more robust than any local solution could be. Cloud providers can undertake security measures such as the encryption of data or regular vulnerability checks, and offer generally tighter maintenance that a smaller company would not be able to take on by itself with its own resources [45, 46].

The composition of its IT Support also influences the risk level of an enterprise. Some small businesses might think they can do without a proper IT specialist in their ranks. However, the presence of someone knowledgeable in that field can be crucial to react quickly and appropriately if some incidents were to occur. On top of other tasks such as providing support and implementing IT infrastructures, they are the best suited to provide quick and adequate responses to a cyber threat [48]. The distinction will be made here between companies that don't have any IT experts among their employees, those that hired one or a few specialists, and finally the ones where an entire department is dedicated to IT.

Vulnerability Identification measures the number of vulnerabilities that a company or at least its IT department is aware of. Despite knowing that there is an exploitable breach, solving the issues and patching the vulnerability might take time. For example, more than a quarter of surveyed enterprise systems were still vulnerable to the WannaCry ransomware in 2020 [47]. Thus, the number of known security vulnerabilities of a company

is relevant in the computation of the risk as it is company-specific information that gives an idea of the magnitude of their issues [33].

The Attack Frequency is also an important attribute to consider. Being the target of relatively more attacks increases the risk of one of them being successful, and not all industries are equally subjected to attacks. A 2020 study analyzed enterprises after re-grouping them into 4 main categories: Business, Healthcare, Government and Education. They found that Governments were less likely to be attacked, whereas the frequency for Businesses and Healthcare facilities was higher [49].

The Attack Frequency concludes the list of chosen attributes to generate data and train the models.

3.2 Dataset Generation

Once the attributes that should be included in the models have been defined, the next step is to generate the training data. As mentioned before, information about an enterprise's cybersecurity is a sensitive topic. Datasets containing such information are not publicly available as they consist of confidential details that companies do not want to share. Thus, a synthetic dataset to train the ML models is generated, using the predetermined attributes.

Once the data is generated, the labels will then be assigned to each datapoint with the mean of an equation that determines the level of risk that the enterprises incur. A total of 100'000 datapoints were generated, further split into 80% of training data and 20% of test data.

3.2.1 Attributes Correlations

To ensure that a realistic dataset is generated, the attributes for one enterprise are not entirely randomly generated. Some conditions are set so that the characteristic of a generated enterprise are coherent. To start, the size of the IT Support depends on the size of the company, hence on the Number of Employees. The Revenue and the Number of Employees are correlated. Cybersecurity Investments are limited for enterprises with low Revenue. Finally, the size of the IT Support department has an influence on the level of Cybersecurity Awareness of the employees.

However, some room for different values is left. These attributes are not totally correlated, and they each bring a contribution to the general equation. This can be verified with a calculation of the Pearson correlation coefficient, which results can be found in Table 3.2.

A positive correlation between some attributes can be observed, but the correlation stays low. The only significantly highly correlated relation is the one between the Number of Employees and the Revenue. This is the case as the constraint linking these two attributes for the data generation is stronger than for other attributes. Although the value might

seem high, it is close to reality: [50] found a similar correlation of 0.885 between the number of employees and the revenue in their study.

Table 3.2: The Pearson correlation coefficient matrix for correlated attributes

Pearson correlation coefficient	Number of Employees	IT Support	Cybersecurity Awareness	Cyber security Investment	Revenue
Number of Employees	1	0.350	0.246	0.421	0.814
IT Support	0.350	1	0.669	0.157	0.283
Cybersecurity Awareness	0.246	0.669	1	0.117	0.201
Cybersecurity Investment	0.421	0.157	0.117	1	0.417
Revenue	0.814	0.283	0.201	0.417	1

3.2.2 Generation of Labels: Risk Equation

The next step is to define the risk probability for the generated data points. After some consideration, the approach of this paper is to use supervised learning models. Unsupervised methods allow the creation of data more freely and would remove the need to generate labels for the data entries. They are however much harder to evaluate, especially with a synthetic dataset. Therefore, the models chosen are using the supervised learning method. To get a prediction, pre-defined labels are needed, and these labels cannot be as freely generated as the attributes. As the ML models use them to learn how to further classify enterprises, they need to be accurate.

To do so, an equation is defined that uses the attributes of each data point and determines whether an enterprise would be at low, medium or high risk to be the victim of a cyberattack. The general idea for the equation is the following:

$$Risk = Attack\ Frequency * (Exposure\ rate - Defense\ rate) + \epsilon$$

To be used in the equation, the attributes have to be normalized and are sometimes only used in combination with one another. The different parts constituting this equation and their utility are explained in more detail in the following paragraphs.

To start, the *Attack frequency* is the attribute Attack Frequency by the industry that has been further normalized.

Then, the *Exposure rate* is composed of three elements: Access Control Rate, Vulnerability Identified and Employee Exposure Rate. The higher the values of the parameters in the exposure rate are, the higher the vulnerability of the enterprise. A value close to zero means that the exposition stemming from that parameter is minimal.

The Access Control Rate (ACR) is calculated from the Access Control Levels (ACL) attribute, which goes from 0 (no access control) to 3 (MFA with 3 different means of authentication). It is normalized with the following equation:

$$ACR = \gamma^{ACL}$$

The graph in Figure 3.1 shows the impact of the transformation on the original data as well as the frequency of each value in the generated dataset. Enterprises that have no authentication, so an extremely weak access control, have an exposure of 1. With one authentication method, the exposure is reduced and for enterprises using MFA, the exposure nears zero.

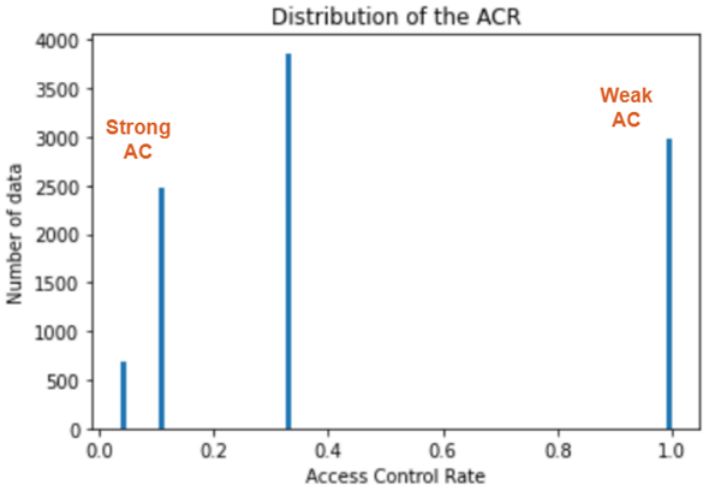


Figure 3.1: Distribution of the data according to the ACR values

For the Vulnerability Identified, the value of Vulnerability Identification which ranges from 0 to 10 is divided by 10 to be normalized between 0 and 1. Finally, the Employee Exposure Rate (EER) is composed of the combination of the attributes of Cybersecurity Awareness and the Number of Employees:

$$EER = \frac{Number\ of\ Employees}{MAX(Number\ of\ Employees)}^{Cybersecurity\ Awareness}$$

Using this equation allows to have the majority of the samples with a high vulnerability. This is the wanted result as there are two categories of enterprises that are more vulnerable: enterprises with a lot of employees and enterprises with a low level of cybersecurity awareness. More employees means more emails, more elements that circulate around the

enterprise and overall more potential vulnerability points. Employees with a low cybersecurity awareness will also bring more vulnerability to the enterprise due to their lack of knowledge and good practices. Only enterprises with a low number of employees that are well trained will get a low exposure from this variable.

This clotures the *Exposure Rate* component. The next part of the equation is the *Defense Rate*. For the Defense Rate, parameters with a high value are the one that contributes to a good defense against cyberattack. Getting a value close to zero shows a lack of readiness regarding that attribute. It is also composed of three different elements: IT Support Level, Cybersecurity Investment Ratio and Data Storage Level.

The IT Support and Data Storage Levels are again just the normalizations of their corresponding attributes between 0 and 1. The Cybersecurity Investment Ratio (CIR) is calculated with the attribute Cybersecurity Investment and Revenue, and put together in the following way:

$$CIR = \frac{Revenue}{MAX(Revenue)} * \frac{CI \text{ - } Adequate\ CI}{Adequate\ CI}$$

The revenue divided by the maximum revenue among all enterprises of the dataset is multiplied by the subtraction of the CI and the Adequate CI, normalized by the Adequate CI - the Adequate CI being a baseline for the adequate percentage that should be invested in cybersecurity. With this equation, the sign of the CIR is determined by the cybersecurity investment: the CIR will be positive if enough money is put in cybersecurity and thus raise the defense rate. It will be negative and lower the defense rate if not enough money is invested. The magnitude of the number is quantified with the revenue. The bigger the revenue, the bigger the issue is when not enough money has been invested in cybersecurity.

Finally, a small error term has been added to add some Gaussian noise.

The complete equation is therefore the following :

$$\begin{aligned} &Risk = \\ &Attack\ Frequency \\ &* [(Access\ Control\ Rate + Vulnerability\ Identified + Employee\ Exposure\ Rate) \\ &- (IT\ Support\ Level + Cybersecurity\ Investment\ Ratio + Data\ Storage\ Level)] \\ &+ \epsilon \end{aligned}$$

The equation was then applied to the synthetic dataset. The following graph displayed in Figure 3.2 shows the frequency distribution of the risk values generated for each of the data points. The labels for each data point are defined according to the calculated value of the risk. A value below 0.1 corresponds to a Low risk, between 0.1 and 0.4 to a Medium risk and above 0.4 to a High risk.

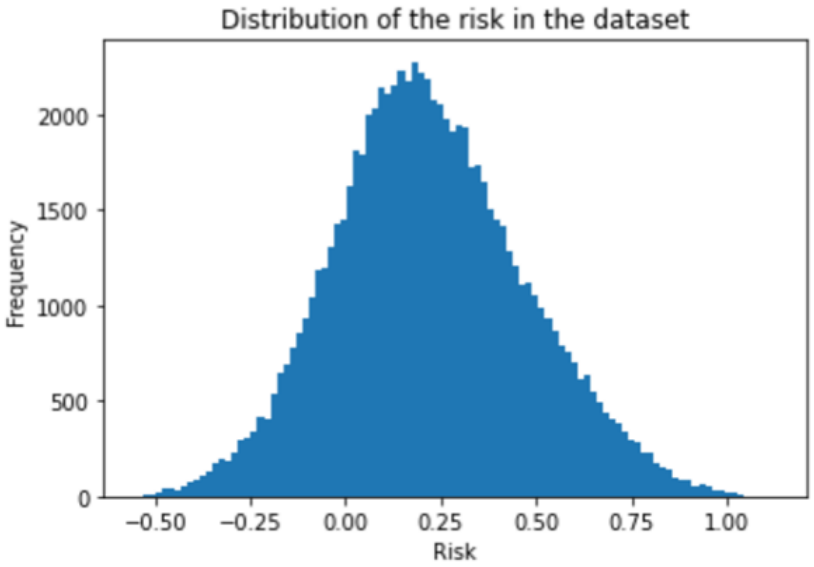


Figure 3.2: Distribution of the Generated Risk Labels

3.2.3 Noise in the Syntethic Dataset

Computing the correlation between the attributes is the first step to ensure a realistic distribution of the values. Another step taken in that direction is the addition of noise in the dataset.

While ideally someone wanting to assess the risk of their enterprise would have all the required information, there might be some companies where it is not the case. Some attributes such as the number of Vulnerabilities Identified or the type of Access Control might not be known, not be assessed correctly or might be misunderstood. Furthermore, mistakes can be made when entering the data. A desirable characteristic of an ML model is to be sufficiently robust enough to give accurate predictions despite some occasional errors in the input.

[63] provides a theoretical perspective of how different types of noise appear in real-world data and affect the ML model prediction. There usually exists two types of uncertainty. One is epistemic uncertainty, which mainly occurs due to the lack of data. In this case, this uncertainty has been highly reduced since 100,000 data points are generated. The other kind of uncertainty is aleatoric uncertainty which originates directly from the process that is being modeled. This uncertainty is the noise in the labels that can not be explained by the inputs. To be more specific, the aleatoric uncertainty can also be divided into two noises, homoscedastic and heteroscedastic noises which are both considered in the data generation process. Homoscedastic noise in this data is the Gaussian noise added during labeling, which has a constant impact. Heteroscedastic noise is the noise in the attributes which changes with the values of the inputs.

Figure 3.3 gives an example of uncertainty disentanglement in toy regression of a sinusoid, produced using an ensemble of 15 neural networks, with standard deviation being computed across the ensemble predictions. Predictive uncertainty is decomposed into aleatoric and epistemic uncertainty, where aleatoric is Gaussian noise added to the data,

and epistemic is higher in out of distribution inputs (indicated by the dashed bars for x $< -\pi$ and x $> \pi$)[64].

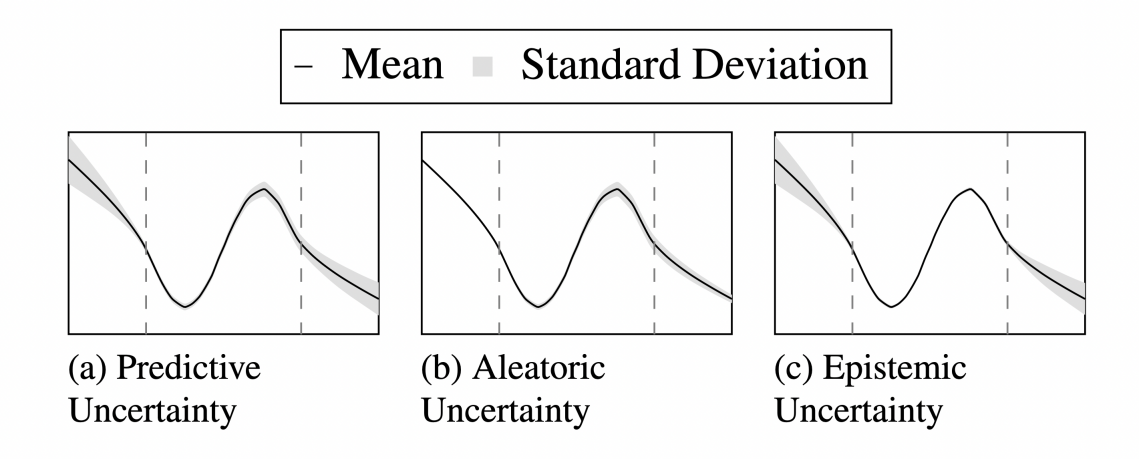


Figure 3.3: Uncertainty Disentanglement

In this regard, the originally generated data is tweaked to create a training dataset including some noise. It is important to note that the original data, the one without noise, is used to generate the labels for the level of risk. This noisy data is only used to train the ML models. The noise is added in the following way: for the categorical attributes, 6.7% of the data points are changed; for the continuous attributes, each data entry gets a different Gaussian noise added. The resulting distributions of the three float attributes are shown in Figure 3.4.

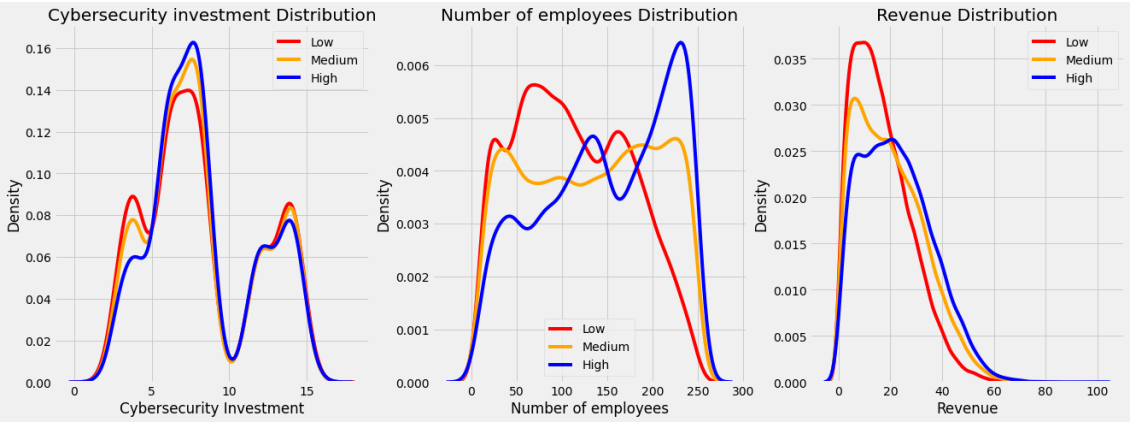


Figure 3.4: Float Attributes Distribution in different Classes

It was mentioned for the Label Generation that some error terms were added in the form of Gaussian noise. Five different noise terms have been tested to find the best version which would lead to better performances from the ML models: no noise, a Gaussian noise standard deviation of 0.01, 0.02, 0.05 and 0.1. As a reference, the result of these five noise additions when inputting the noisy data in the generating equation can be observed in Figure 3.5. As the equation is here being used to label the data points, the end noise term is set to zero so as to not influence the classification). Unsurprisingly, the more noise added the less accurate the equation prediction. It can be noticed that the accuracy goes

down faster than linearly. A standard deviation of 0.1 results in an accuracy of around 75%, against around 84% for a 0.05 standard deviation and 91% without any added noise.

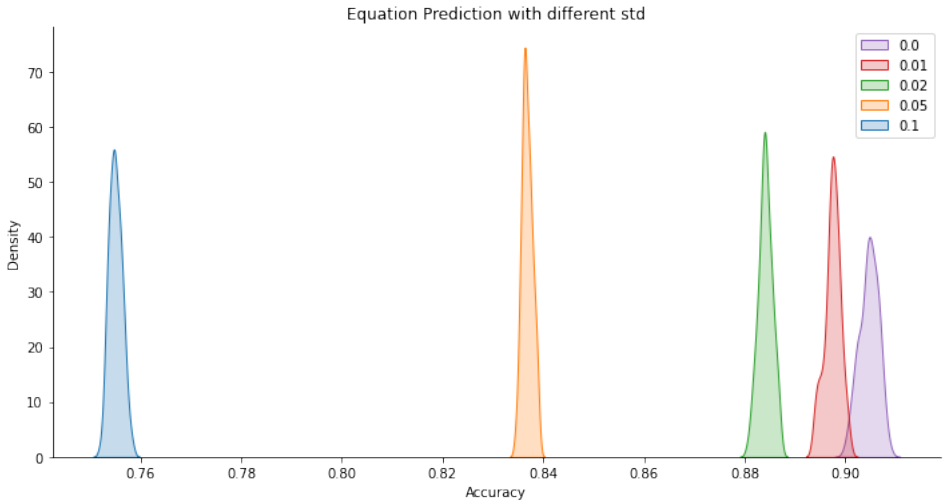


Figure 3.5: Accuracy Distribution using the Equation with different Noise Terms

The addition of noise, both directly in the data points as well as in the equation, serves two purposes. First, it allows approximating what real-world data would be, containing incoherence and small errors difficultly noticeable. Moreover, it also trains the ML models to be better than the equation that generates the risk. One issue with creating an equation to generate labels is that this equation has perfect accuracy: if it is used to create the labels, it could also be used to classify new data. The equation would however be highly impacted by some wrong data as it cannot be prepared for this. Training the models with noisy data that have the correctly assigned labels lead to more robustness, and a better chance to get an accurate classification with some slightly wrong inputted data.

3.2.4 Missing Attributes

Another advantage that the trained models have over a linear equation is the adaptation to missing values. As aforementioned, some enterprises might have issues properly assessing some of the values needed, which would lead to erroneous data. In the worst-case scenario, the number might be totally unknown and not only slightly incorrectly estimated, which leads to complications.

Among the nine attributes that are initially considered to predict the risk, two attributes were identified as being potential issues. These two attributes are the Cybersecurity Awareness and the Vulnerability Identification. The level of cybersecurity awareness of the employees of an enterprise can be challenging to assess as there are no fixed metrics to measure it, and the value set for this attribute depends heavily on subjective opinions. On the other side, the number of already identified vulnerabilities is a more technical attribute that some companies might not be aware of if they do not regularly run such tests.

Either one - or both - of these attributes can therefore be left out, and the models are adapted to deal with such cases. The other seven attributes however are considered sufficiently objective and attainable information. They are therefore mandatory inputs that have to be provided to get a result from the ML models.

3.3 Machine Learning Models Training

The initial data, without any added noise, was used for the labeling. It was then slightly distorted with the addition of some noise in the data points. The following section focuses on the training of the ML Models to predict the labels on test data. As defined earlier, the ML Models that are used are RF, SVM and NN. The training and testing are done on the noisy data. For each model, four different classifiers are trained. The main one is a normal classifier with all the dataset attributes. The other three are variations around the potentially missing attributes.

There are different ways of dealing with missing data. A first option is to input some value to the missing data, usually the mean or the median of the values that this attribute can take. A second option is to entirely get rid of that attribute. To be able to predict labels with an attribute missing, another model without this attribute has to be trained. Between these two options, removing the attribute altogether showed better results for all three ML models. The accuracy of the models was diminished, but not as much as with the inputted values.

Thus, three other models are trained alongside the main one for each of the ML methods: one without the Vulnerability Identification attribute, one without the Cybersecurity Awareness attribute and one without both. There are therefore in total 12 models that are fitted and can be used to predict the risk from new data.

The first model used is the capable Random Forest Classifier in Scikit-Learn. This classifier is not expected to give a spectacular performance, but it establishes a baseline for further results. The important parameters of RF are selected as follows in Table 3.3. Theoretically, given enough data, the larger the number of estimators that are added to the model, the more accurate the prediction will be. However, considering the training speed, it is set to 500.

Table 3.3: Initial RF Parameters

Data Preprocessing	Number of Trees	Maximum Depth	Class Weight
False	500	Pure Leaf	Balanced

For the SVM, Data Normalization is necessary to keep the attributes unbiased. The SVC method in Scikit-Learn was selected since it can be trained much faster than the NuSVC method and in the meantime also solve non-linear classification problems. The RBF kernel

is the most powerful kernel among the kernel functions. it was therefore decided to use the RBF kernel with coefficient $\gamma = \frac{1}{|feature|}$ and regularization term $C = 20$.

For the NN, different combinations of parameters have been tested through GridSearch to reach the best accuracy. The best parameters to train the complete model were found to be a model using the ReLU activation function with 4 hidden layers, with the Adam optimizer and an adaptive learning rate of 0.0001.

3.4 Front End

To interact with the aforementioned models, a front end was developed. In this chapter, the implementation of the platform is presented. The first part will show the techniques and details, including user scenarios and interaction for the front end. The second part will detail the logic of the algorithms, that is the SVM, NN and RF, as well as the API calls in the back end. The main view of the front end is presented as Figure 3.6, where the user performs the prediction and receives the results. The source code is publicly available on GitHub.

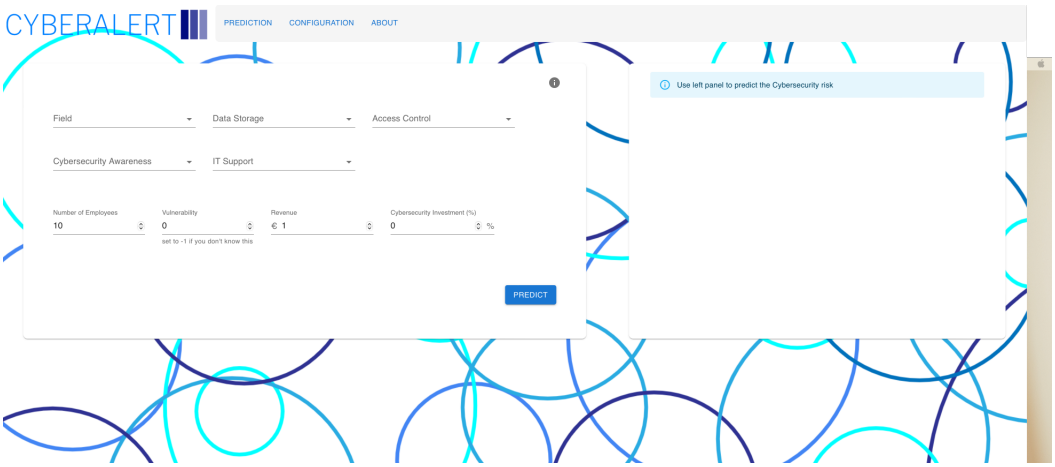


Figure 3.6: Overview of the main page

Figure 3.7 gives a high-level overview of the interactions between the user and the different components of CyberAlert. The flow is triggered by the user interacting with the main web view. The user inputs the value of the attributes in the predicting view. With a click of the predict button, a request is sent to the back end. The API components in the back end handle the incoming request and transfer the data from JSON to an array-like data structure. The ML classifiers then take the requested data and, at the same time, use the pre-trained classifiers to perform the prediction based on the different parameters contained in the requested data. Then, the API component transforms the output again to a JSON datatype and finally returns the result to the front end by a response. If there is an error happening in the back end, the back end sends an HTTP error message. The front end catches the error message and displays it to the user.

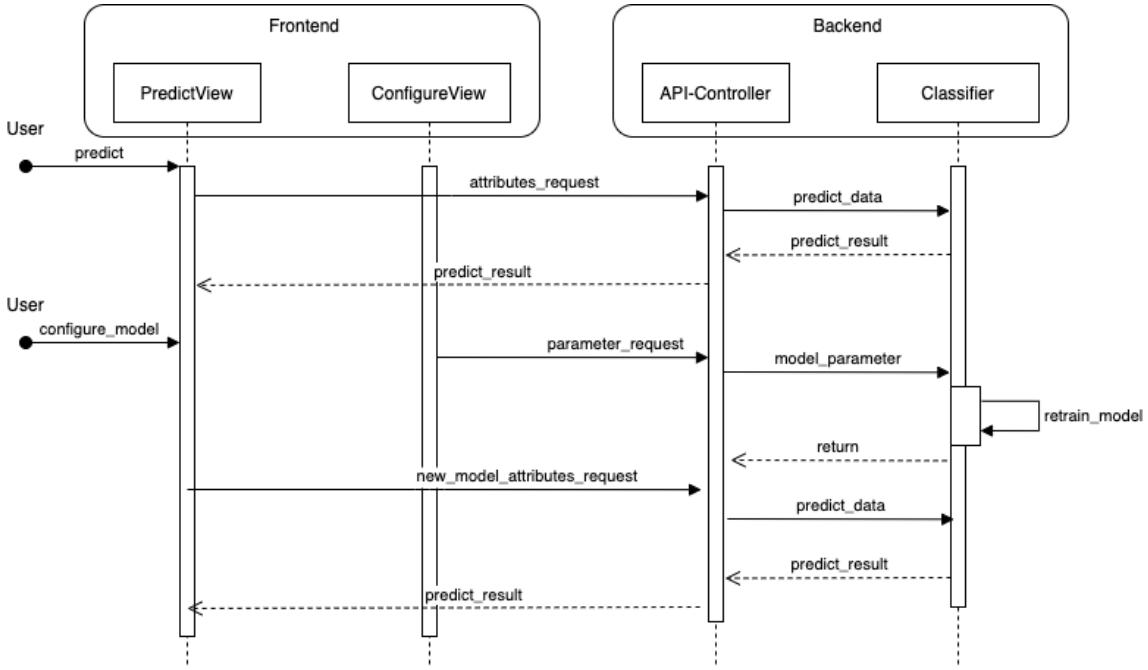


Figure 3.7: Sequence diagram of CyberAlert

3.4.1 React

The front end of the CyberAlert platform is implemented using React [65], a JavaScript library for building user interfaces. There are several advantages to choosing React as the frontend library. First, it is easy to install and start a small project. When learning how to use React, the components and concepts are relatively simple to pick up, which means the learning curve is not as high as in other frameworks. Secondly, the components in React make it more efficient to build the user interface, as the same elements can be used multiple times. In addition, React has huge community resources. Therefore, it is much easier to build a high-quality User Interface (UI) with the support of community resources like text fields, buttons, sliders, etc. This allows React users to focus only on the content of their project instead of writing the interface elements repeatedly. The concept of virtual Document Object Models (DOM) also plays an important part in React. DOM is a tree-like interface that represents HTML and XML code. A web browser will have a DOM model to render output. It shows the HTML code as a node in the DOM tree. By virtualizing and keeping DOM in memory, React is capable of fast rendering and all the view changes can be prepared in the virtual DOM. The specialized algorithm and virtual DOM states in React will calculate the most efficient way to apply recent view changes without sending updates in a frequent way. With that small change in the DOM, it can have a minimum number of updates to make the browser updated quickly and efficiently, which results in an overall performance boost.

3.4.2 MUI

When considering user interface, a diverse set of interactive elements for users to interact with is often necessary. It is time-consuming to develop the UI from scratch, especially without proper knowledge of UI design, as it can get troublesome and frustrating.

Material UI (MUI) [67] is used extensively in this project for this matter. MUI is a library of UI components that can be used to build React applications. After installing the package, the large reusable component library can be accessed, including the buttons, sliders and icons. And the components are aligned with Material Design by Google by default. Because of this, the MUI components will probably look familiar to the potential user. The default styling already has a high standard and only needs some slight adjustments and tuning, as can be seen for example in Figure 3.8. Using MUI also ensures that UI is consistent with the design language. The reusable elements can significantly reduce the number of errors and workload.

Building the UI components from scratch would increase the complexity of the design workload. One would need to verify that there are different formats for the same design. For example, the alert message may only shows an icon rather than the icon with text if they appear in a tighter space. However, using the library provided by MUI, which is optimized for the customization of the components, helps to flexibly adjust to different scenarios. When the default state of an element is not what is needed, it can always be restyled and easily transformed to suit the project better. After the implementation of the front end, the MUI can also be of help with the maintenance and documentation. The well-designed API documentation from MUI can be used as a reference to develop CyberAlert's documentation following a similar standard.

```
<FormControl variant="standard" sx={{ m: 2, width: "30ch" }}>
  <InputLabel id="demo-simple-select-label">
    Cybersecurity Awareness
  </InputLabel>
  <Select
    labelId="demo-simple-select-label"
    id="demo-simple-select"
    value={awareness}
    label="Cybersecurity Awareness"
    onChange={(event: SelectChangeEvent) =>
      setAwareness(event.target.value)
    }
  >
    <MenuItem value={0}>Low Awareness</MenuItem>
    <MenuItem value={1}>Moderate Awareness</MenuItem>
    <MenuItem value={2}>High Awareness</MenuItem>
    <MenuItem value={3}>Very High Awareness</MenuItem>
    <MenuItem value={-1}>I dont know</MenuItem>
  </Select>
</FormControl>
```

Figure 3.8: Applying MUI in front end development

3.4.3 AXIOS

When dealing with the web application in React, the most common task is the communication between the front end and the back end. Usually, XML, HTTP Request and Fetch API help to complete such requests. For the implementation of this tool, the Axios library is used to make communication easier. Axios is a JavaScript library for making HTTP requests based on the Promise API [66]. It has a succinct code with only three clauses to send the request data into JSON and get the returned data in JSON in a single response. It is also relatively easy to handle errors with Axios, as shown in Figure 3.9. Unlike Fetch, which does not notify of the rejection of the promise, Axios throws network errors. If there is a bad response (e.g. Error 404), the request is rejected, and Axios returns an error that can be checked to identify the type of error.

```
axios
  .put(url: "http://127.0.0.1:5000/modif", data: { param: param, attr: newAttributes })
  .then(onfulfilled: function (response) {
    props.updateModifResult(defaultResult: {
      result: response.data.result,
      mode: response.data.mode,
      model: response.data.model,
      active: true,
    });
    console.log(message: response.data);
    setModLoading(value: false);
  })
  .catch(onrejected: function (error) {
    console.log(message: error);
  });
```

Figure 3.9: Applying Axios in front end development

3.4.4 Web-based Interface

After implementing different views and user interfaces, the web-based tool is finally ready for users. There are three views: the Prediction page, the Configuration page and the About page. The primary view is the Prediction page. When users land on the project page, they first see the Prediction page. On the left, they can input the attributes of the company for which they want to predict the risk (Figure 3.10). The Field, Data Storage and Access Control are on the first row. Cybersecurity Awareness and IT Support are on the second row, and the Number of Employees, Vulnerability, Revenue and Cybersecurity Investment is on the third row.

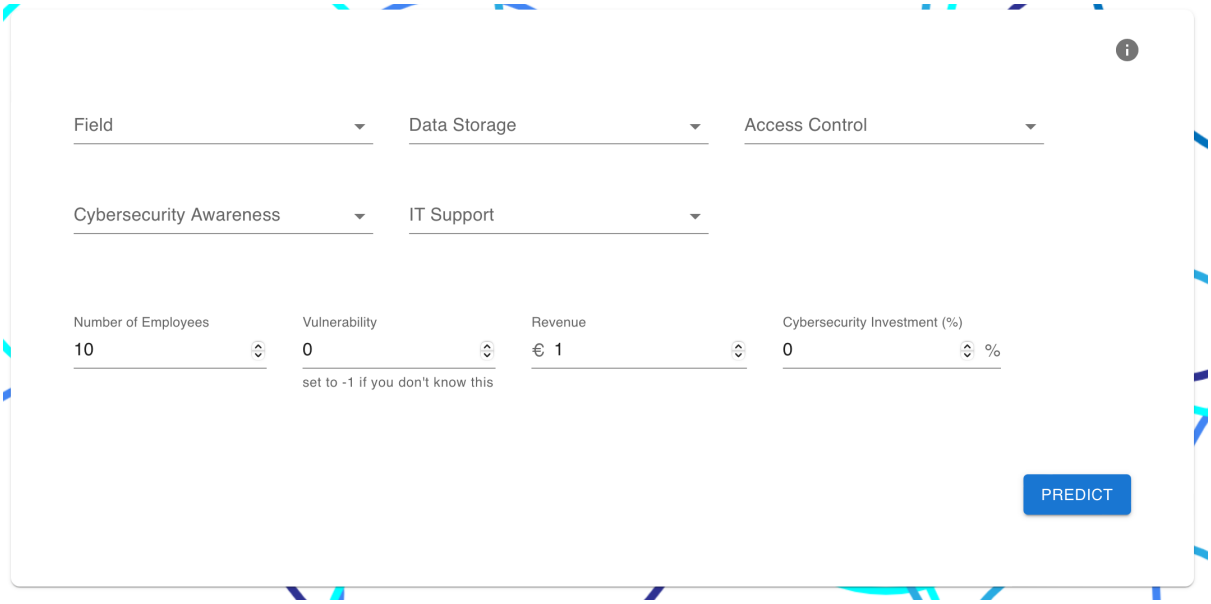


Figure 3.10: Attributes input panel

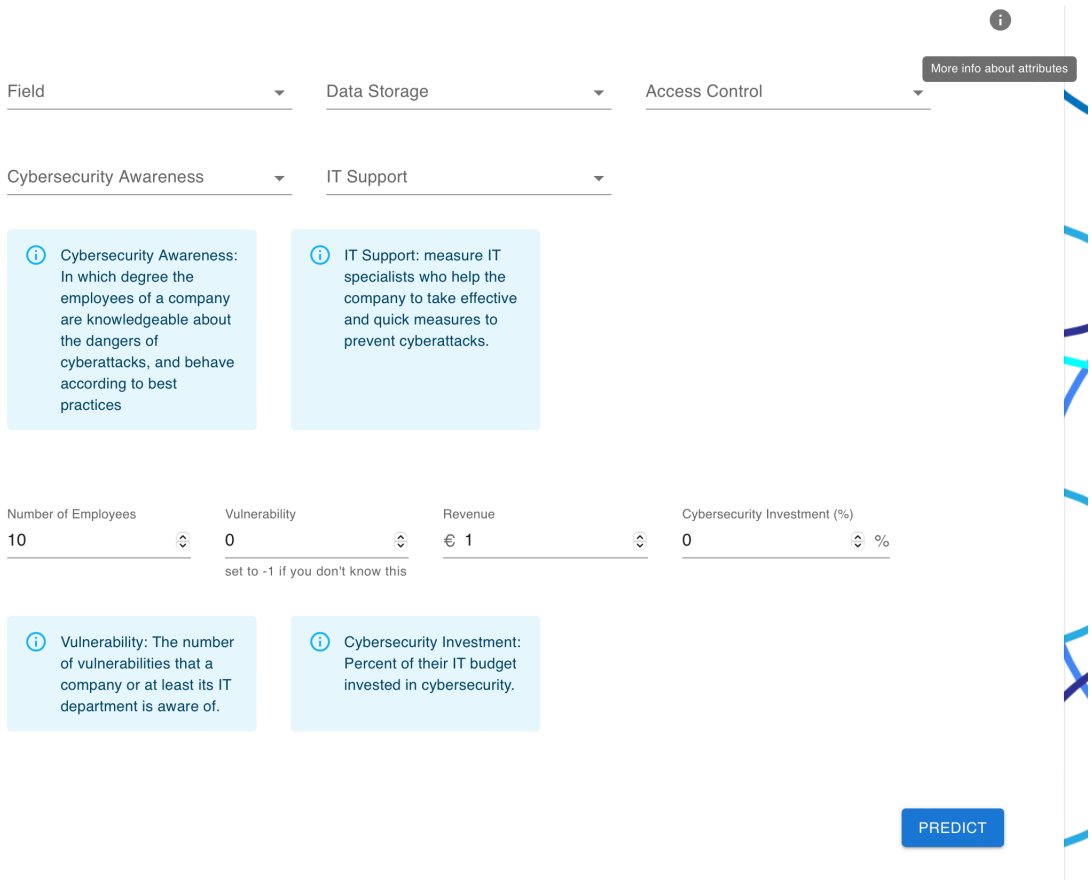


Figure 3.11: Attributes information

If it is their first time using the platform, the users may need clarification about the meaning of the different attributes. They might also need help knowing what to enter in

each field. There is an information icon on the top right of the predicting panel. When the user clicks on that icon, a detailed explanation is shown (Figure 3.11). For example, the user may need clarification on what percentage the Cybersecurity Investment field refers to. The explanation will guide them to enter the percentage of their IT budget invested in cybersecurity.

After entering the attributes of the company for which they want to know the risk, the user can click on the "Predict" button to get the final prediction based on the ML models. This result is shown to the user on the Result panel on the right, which is divided into three parts. All three models are used to predict the risk. Therefore, there is a place for displaying the results from the SVM, NN and RF, aligned vertically from top to bottom, as shown in Figure 3.12. The result is categorized into three different severity: Low, Medium and High. When the front end gets the numerical response from the back end, it is transformed into the corresponding string type severity level and shown to the user.

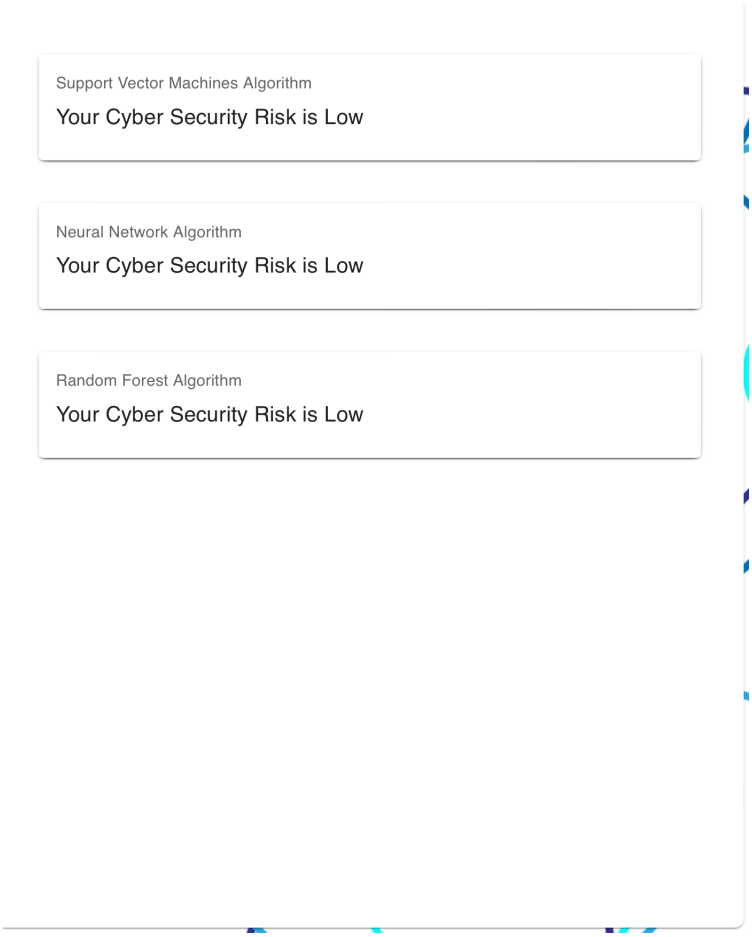


Figure 3.12: Panel presenting the results from three algorithms

Another view that the tool provides is the Configuration page. The user can access the Configuration page through the navigation bar at the top of the screen. On this page, users can take a deeper look at the models used for prediction. It offers the option for users to customize the models. By comparing the results after several modifications, they can observe the impact of each parameter on the model. If the user is interested in how

they can make the model more accurate and fit their company, this is where they have the opportunity to modify the parameters for each ML model. The user can choose which algorithm they want to modify on the selection menu on the top, as shown in Fig 3.13.

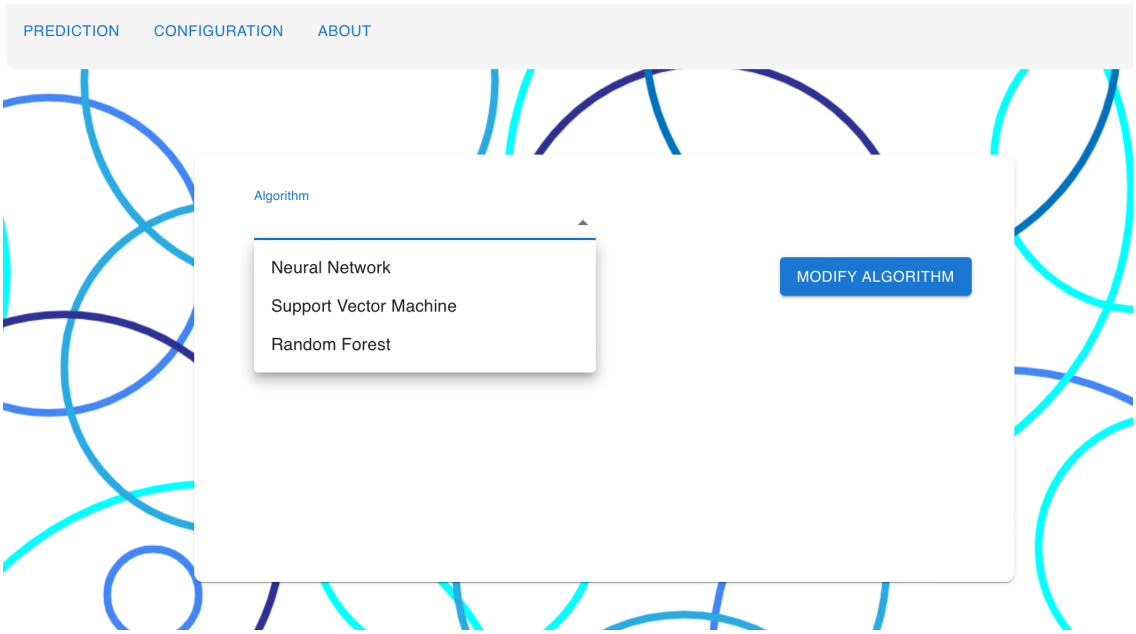


Figure 3.13: Configuration panel to modify the parameters of the algorithms

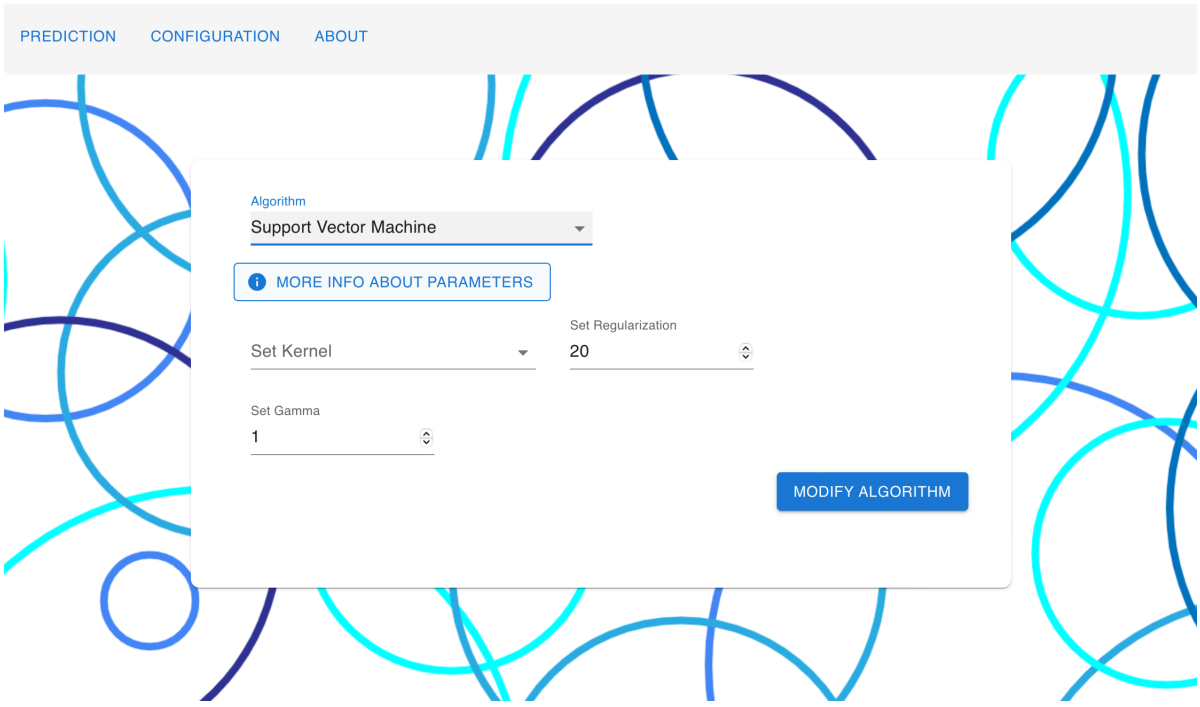


Figure 3.14: Configuration panel for SVM

Figure 3.14 shows the example of a user who is expecting to modify the parameters for the SVM. Once the desired algorithm is selected, they can change the kernel from Linear

to Poly. They can also set the regularization to a different number of their choice and finally tune the gamma number to optimize the SVM model.

If the user is not familiar with the parameters on the Configuration page, they can get additional explanations by clicking on the ”More Information” button, just like on the Prediction page. For each of the parameters, it is explained how the parameter works or the roles that the parameter plays in the model in detail (Figure 3.15). This provides the user with a better experience as they are supported in the use of the tool.

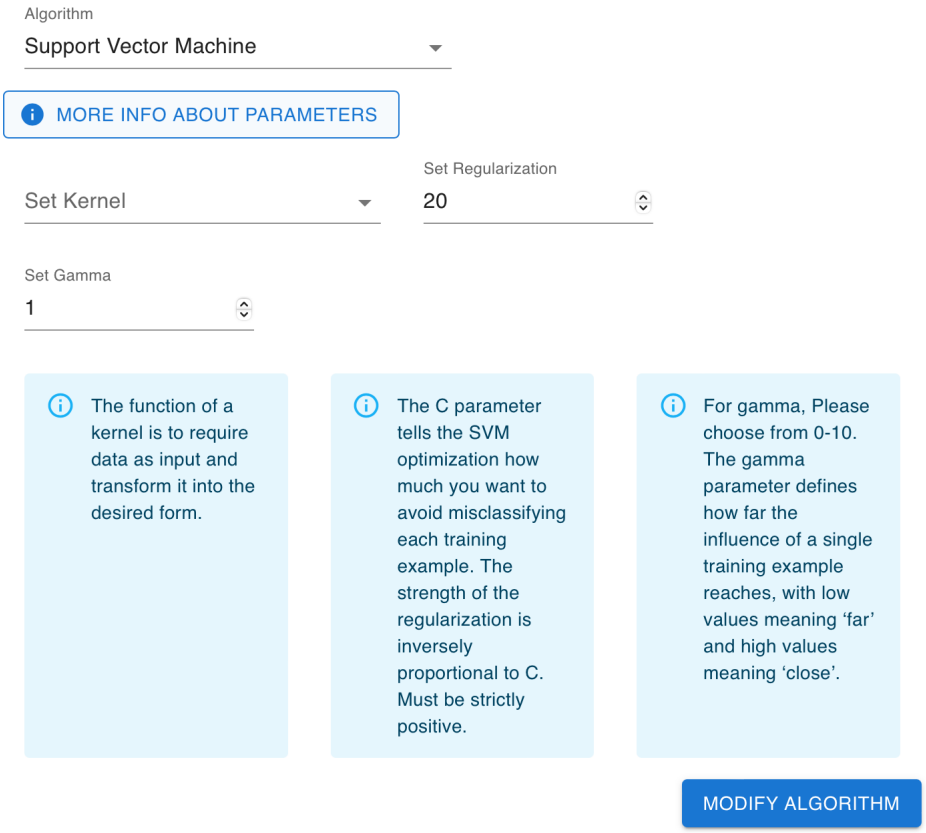


Figure 3.15: Detailed explanation of the parameters for SVM

The same setup is applied to all the models, as can be seen in Figure 3.16 and Figure 3.17. For the NN model, the user can change the activation function from ReLU to Tanh, Logistic or Identity and set the Hidden Layers and Learning Rate. For the RF model, the user can set the estimator, depth and minimum sample split. All the parameters mentioned above have a corresponding explanation on the Configuration page to make sure that the modification effectuated by the user is aligned with their intentions.

Once they have changed the configurations, users can save their choice by clicking on the ”Modify Algorithm” button on the bottom right. If the parameters are successfully saved, a banner will show up to notify the user that their configuration was accepted. After that, the user can go back to the Prediction page. Once the parameters are saved for a modified model, a new button will appear at the bottom of the attributes panel as

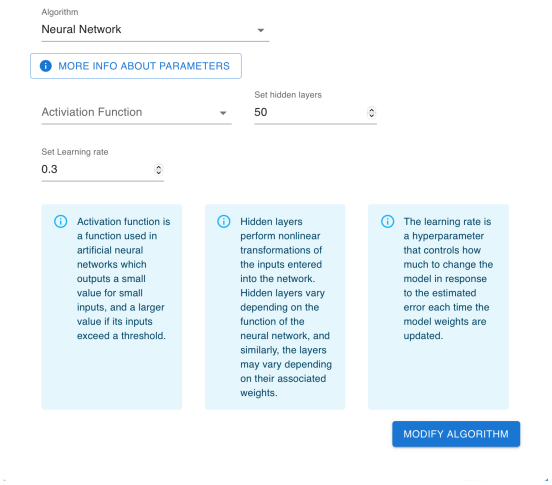


Figure 3.16: Detailed explanation of the parameters for NN

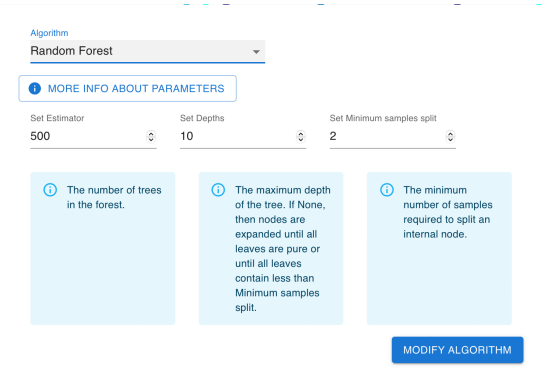


Figure 3.17: Detailed explanation of the parameters for RF

shown in Figure 3.18. Users can use that new button to obtain a prediction using their personalized configurations. Since the platform needs to train the modified model again, the waiting time could vary depending on the parameters and the type of model. The "Predict using new model" button will show a loading circle to inform the user the system is still running. Once the training and the predictions are completed, the result will be displayed on the right panel with the helper text labelling the result from the modified model, as shown in Figure 3.19.

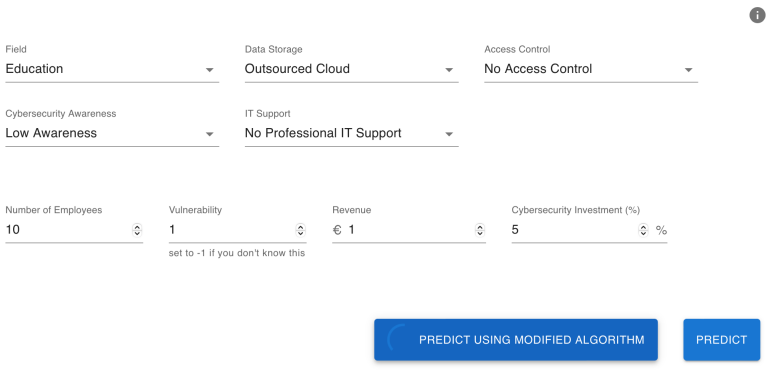


Figure 3.18: Prediction with the modified algorithm

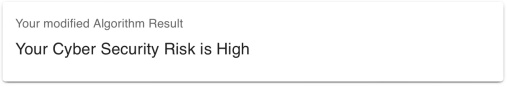


Figure 3.19: Result from the modified algorithm

The last page of CyberAlert is the About page, displayed in Figure 3.20. There, users can find more information about the project at the very top of the page, followed by information concerning the team involved. On this page, the user is also provided with more information about the attributes used for the prediction. In the detail section, every attribute required for prediction is listed with an explanation concerning the reason for including such an attribute. Users can refer to this About page in case they have questions about the selection of attributes. At last, the GitHub link of this project and the report link are included at the bottom of the page for anyone interested in this project.

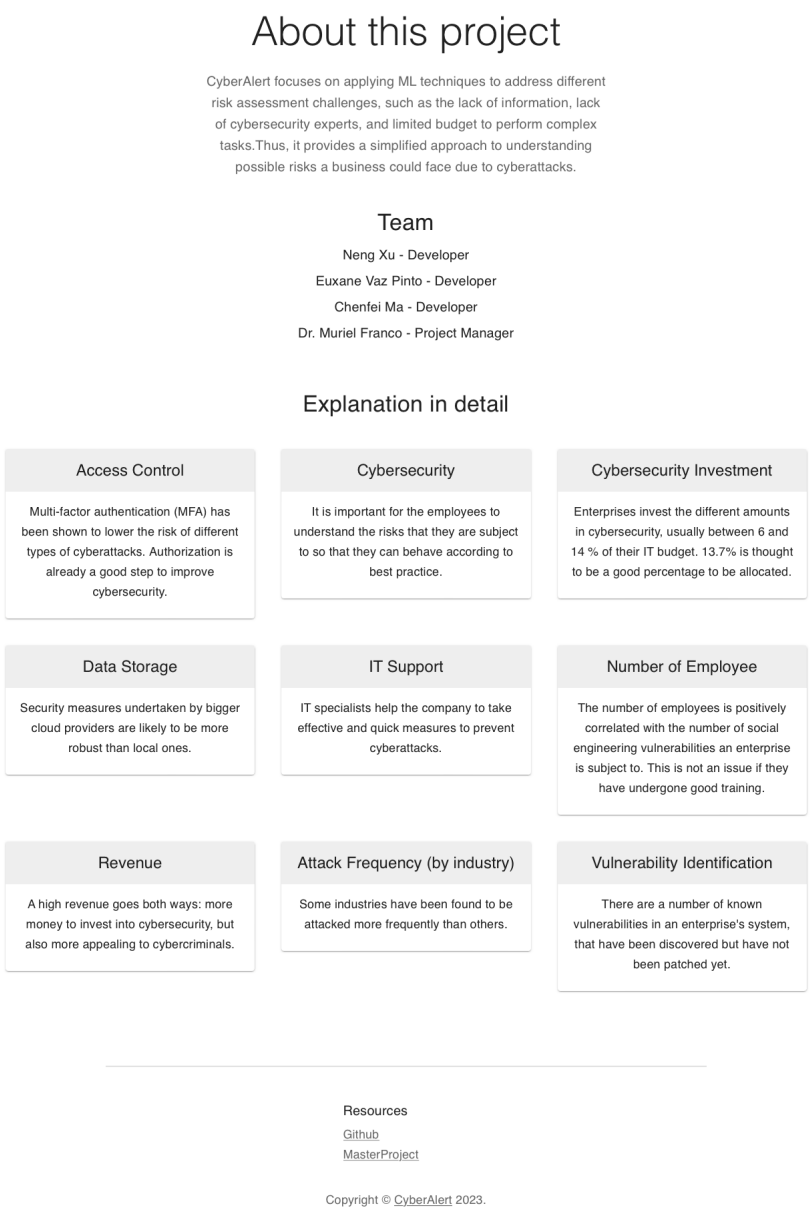


Figure 3.20: The About Page

Chapter 4

Model Evaluation

This chapter presents the evaluation of the three implemented ML models (*i.e.,* NN, SVM, and RF). They are evaluated through an accuracy comparison, followed by an analysis of their confusion matrices and F1-score. Next, a visualization of the classification process and specific measurements based on each model are also provided for a better understanding of the behaviors and performances of the implemented models.

4.1 General Methods

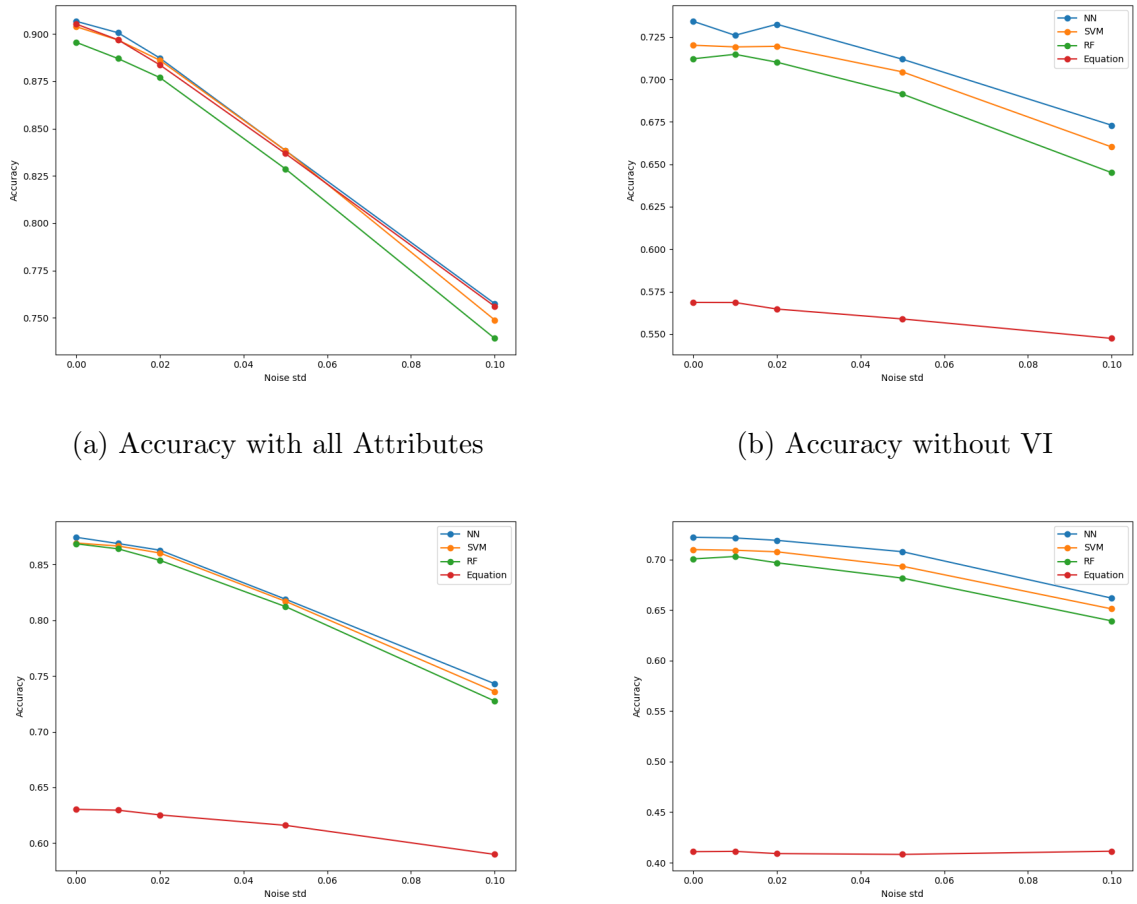
The general evaluation methods for supervised classification models are implemented to measure the performance of the three different models. The confusion matrices and optimized F1-score will be the focus of this section. Each of the three models was evaluated in the four different modes previously defined: *(i)* with the full attributes (All), *(ii)* without the Vulnerability Identification attribute (NoVI), *(iii)* without the employee's Cybersecurity Awareness (NoAW), and *(iv)* without both of these attributes (noBOTH).

The equation that was implemented to create the labels could also be used for the prediction. Therefore, the accuracy of the equation for the different modes was also computed as an object of reference to see if there is an actual improvement in performance by using ML models over the original labeling equation. When the equation is used to predict the labels, the additional noise term is set to zero.

The Gaussian noise added to the equation to generate the labels was a parameter that had to be chosen. In order to find the most fitting noise to add, multiple options ranging between no noise and Gaussian noise with a standard deviation of 0.15 were tested. The results were then compared to find out which one was the most meaningful. The final label Gaussian noise standard deviation this project added in the synthetic datasets is 0.05 based on the model performance.

4.1.1 Performance Evaluation: Accuracy and Confusion Matrix

Accuracy (ACC) is calculated as the number of all correct predictions divided by the total number of datapoints. Figure 4.1 shows the accuracy of the models for the four predetermined modes. The accuracy of the initial equation is also represented to assess whether the models can outperform the initial labeling equation.



(a) Accuracy with all Attributes

(b) Accuracy without VI

(c) Accuracy without AW

(d) Accuracy without VI and AW

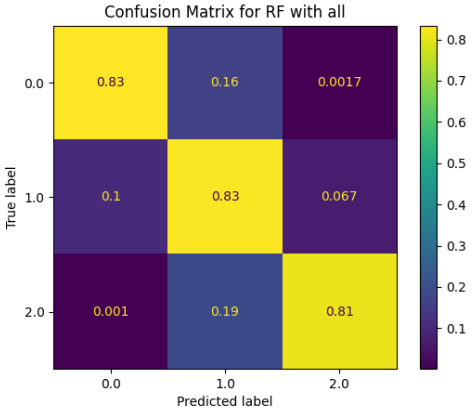
Figure 4.1: Accuracy of the four ML models with added Gaussian Noise

The first observation that stands out in this graph is the important impact of removing some attributes on the equation. The accuracy of the equation drops to around 50% and is largely outperformed by the ML models. On the other hand, all three ML models are able to overcome the loss of one or more attributes. Vulnerability Identification (Figure 4.1c) is the attribute that impacts the accuracy the most, making even the noiseless data's accuracy drop from 90% to 70%. The impact of Cybersecurity Awareness is much weaker, as shown in Figure 4.1d. The removal of that attribute only reduces to around 85% the accuracy of the noiseless data. In addition, although this noiseless accuracy is significantly lower than the one of the complete model with all attributes, both of these modes reach a similar accuracy of around 75% when a large noise is added. This shows

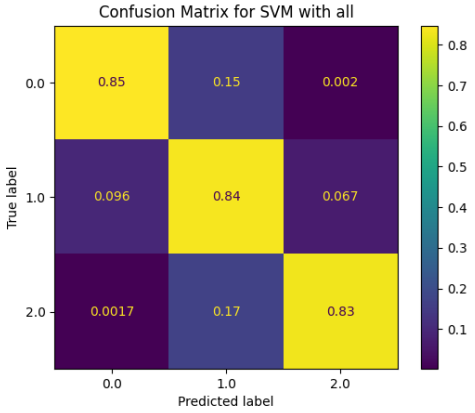
that the cybersecurity awareness attribute is not one of the most important for any of the three ML models.

Another element of importance is that the NN is the only model that consistently outperforms the result of the equation in the complete model (Figure 4.1a). The difference in accuracy across the different noises stays relatively similar between both of these models, the NN being only slightly more accurate than the equation. On the other hand, the SVM algorithm shows better results for a lighter noise but gets outperformed by the equation when a bigger noise term is added. The RF algorithm is consistently performing the worse, always being marginally less accurate than the other models.

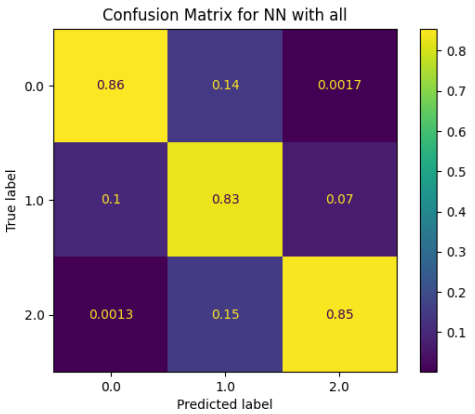
Finally, the four graphs can be used to determine which error noise would be the more adequate to keep for the final model. To keep a balance between the need to have noisy data and to try to get closer to real-world data without overly deteriorating the accuracy, the standard deviation of 0.05 has been chosen. This decision is further consolidated by the results of the F1 score which will be discussed in the next section.



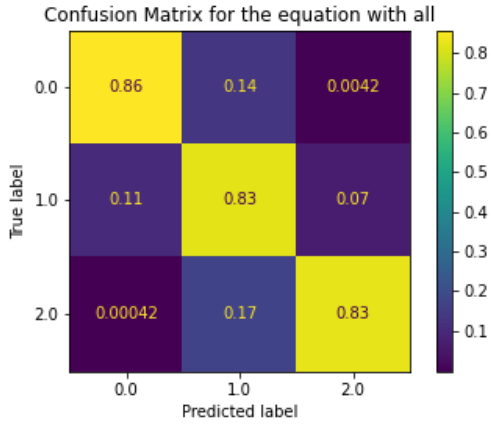
(a) Confusion Matrix of the RF



(b) Confusion Matrix of the SVM



(c) Confusion Matrix for the NN



(d) Confusion Matrix for Equation

Figure 4.2: Confusion Matrix using all features, with noise std = 0.05

In the field of machine learning and specifically multi-classification problems, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm of supervised learning, which is seen as a general method for model evaluation.

As the most adequate standard deviation of the noise was selected to be 0.05, it is therefore with the data that was labeled using this noise that the confusion matrices were computed to compare the performance of the different ML models.

Confusion matrices allow seeing more precisely the proportion of data points from each class that was mistakenly assigned to the wrong label. Figure 4.2 shows the confusion matrices generated for each ML algorithm with all attributes considered. The value 0.0 corresponds to a Low risk, 1.0 to a Medium risk, and 2.0 to a High risk. The accuracy of the correctly assigned labels for the three classes is pretty evenly distributed among the four models. The "Low risk" label is the one that is consistently the more correctly assigned, with an accuracy ranging from 83 to 86 percent.

In addition, the mislabelled data is most often mislabelled to the neighboring class - there are more "High risk" data points that are predicted as "Medium risk" than "Low risk", which is a good feature of each of the models. In the case of mislabelling, the most problematic error in labeling is confusing a "High risk" with "Medium" or "Low risk". This would give a sense of security to an enterprise that is actually very vulnerable to cyberattacks. The equation (Figure 4.2d) has the lowest proportion of "High risk" data points labeled as "Low risk", only 0.42% of the data. However, the NN (Figure 4.2c) gives the best "High risk" accuracy overall with only 0.15 of the data being labeled incorrectly, mainly as "Medium risk". In line with the accuracy evaluation, the RF (Figure 4.2a) performs the worse out of the four, not reaching higher than 83% of accuracy for any of the labels.

4.1.2 Adapted F1 Score

Only comparing the accuracy is not enough for the model evaluation, since in business scenarios, most data will not be balanced. The F1 score, defined as the harmonic mean of precision (the number of true positives over the positive predictions) and recall (the number of the true positives over the actual positive labels), reflects more accurate model information when the classification problems have unbalanced datasets. The original F1-score for the multi-classification problem is computed as follows:

$$F_1^i = \frac{2}{recall^{-1} + precision^{-1}}, \tag{4.1}$$

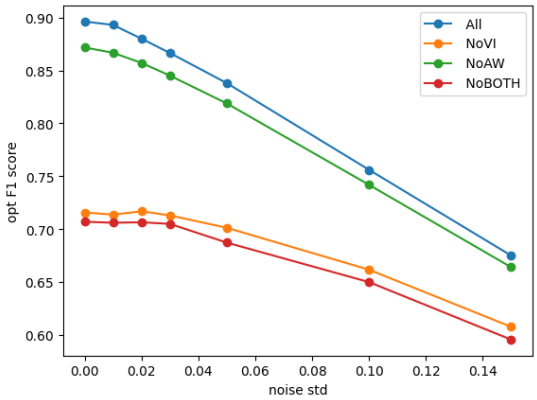
where the F_1^i means the F_1 score for class i, i$\in \{L(Low), M(Medium), H(High)\}$. This paper follows the idea of the Weighted F1 score, assigning different weights to the F1 score of different classes. In this case, the true label 'High' should weigh more as it brings more danger to the company if the predicted label is 'Medium' or 'Low'. The adapted F_1^{adp} score is computed as:

$$F_1^{adp} = 0.2 * (F_1^L + F_1^M) + 0.6 * F_1^H \tag{4.2}$$

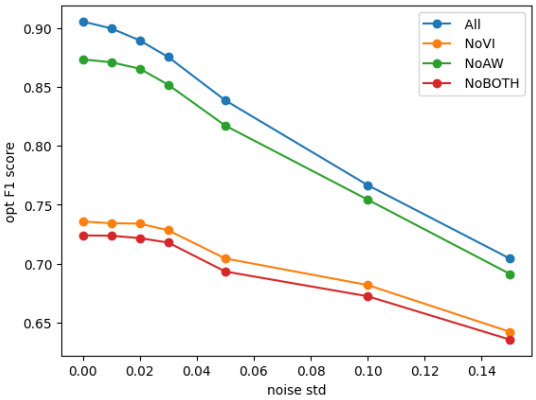
Table 4.1: Adapted F_1 score for each ML models with 0.05 noise

	SVM	**RF**	**NN**
All	0.843	0.834	0.850
NoVI	0.721	0.701	0.712
NoAW	0.827	0.818	0.836
NoBOTH	0.710	0.691	0.713

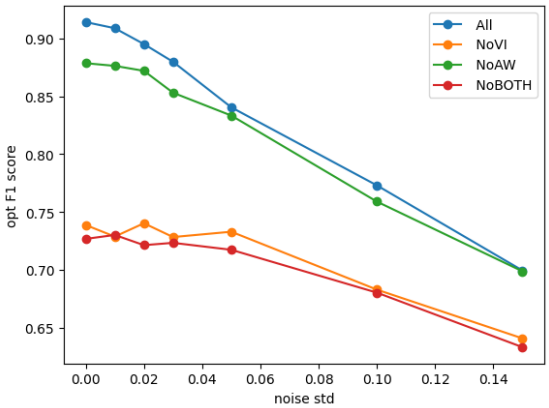
In Table 4.1, the optimized F1 score is computed for all three models. Similar to accuracy, the RF model provides the worst performance in all four modes. The SVM and NN both achieved a high F_1 score. The difference is that SVM outperforms NN with no VI attribute compared to the accuracy result. The confusion matrices in Figure 4.2 also lead to the conclusion that all three ML models succeed to predict decently well on biased datasets and specific labels since the ratios on the four corners of the confusion matrices are low.



(a) F1 score for RF



(b) F1 score for SVM



(c) F1 score for NN

Figure 4.3: F1 score for all three models in different noise term

Figure 4.3, on the other hand, focuses on how the different noises impact the F1 score for all three models. The larger variance added to the predictions of synthetic data, the worse performance of the models. The impact is nearly-linear but there is a drop point around the std = 0.05, especially for the NoAW and NoBoth modes. This observation supports the decision of choosing a standard deviation of 0.05 as the final noise addition.

4.2 Targeted Methods

This section discusses the specific evaluation methods based on the different properties of the models. Despite performing the worst among the three black-box models, the RF algorithm provides the most interpretive insight into the data and the process of training.

The feature importance of RF is computed by the model built-in Gini importance. The features for internal nodes are selected with gini impurity. For each feature the decrease of the impurity is collected and averaged over all trees. Figure 4.4 explicitly shows the feature importance for RF in the synthetic dataset. The Number of Employees and Vulnerability Identification are the leading terms in the model prediction, which also shows the rationality of the data generation. More employees correspond to the exponential complexity of company information exchange, and Vulnerability Identification is a straightforward attribute that affects the final labeling.

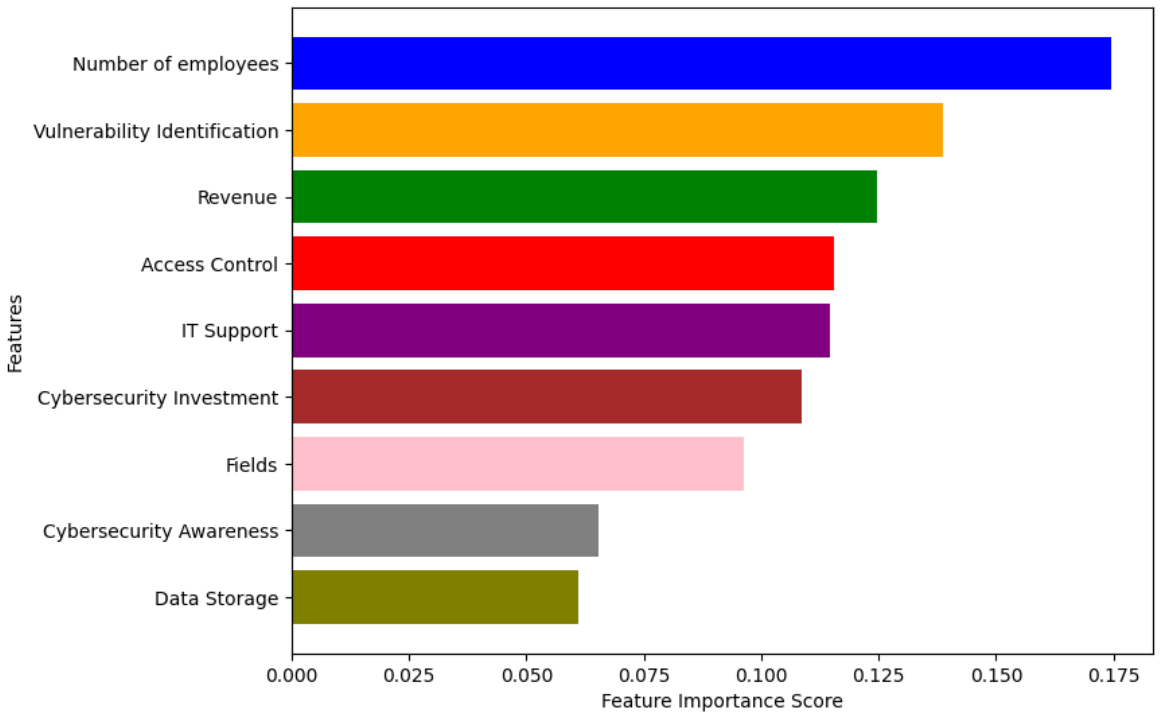


Figure 4.4: The feature importance score generated by RF

It is interesting to point out that this ranking is in alignment with the findings from the accuracy and F1 score of the models without Vulnerability Identification. As this

attribute has a strong impact on the prediction, removing it leads to more mislabeling. The results which omit the Cybersecurity Awareness, on the other hand, maintained a higher accuracy. This correlates with that attribute being the second least important variable. Besides, another observation is that there are no redundant features in the data generated since the importance score of all attributes is not low.

The RF training process can also be visualized, in this case by selecting the trees in it. Figure 4.5 shows one of the estimators (trees) in the RF. The left nodes of the subtree are always the true assignment to the specific class. "gini" is the variance of labeling to different classes. Access Control is the leading attribute for this tree, followed by Vulnerability Identification.

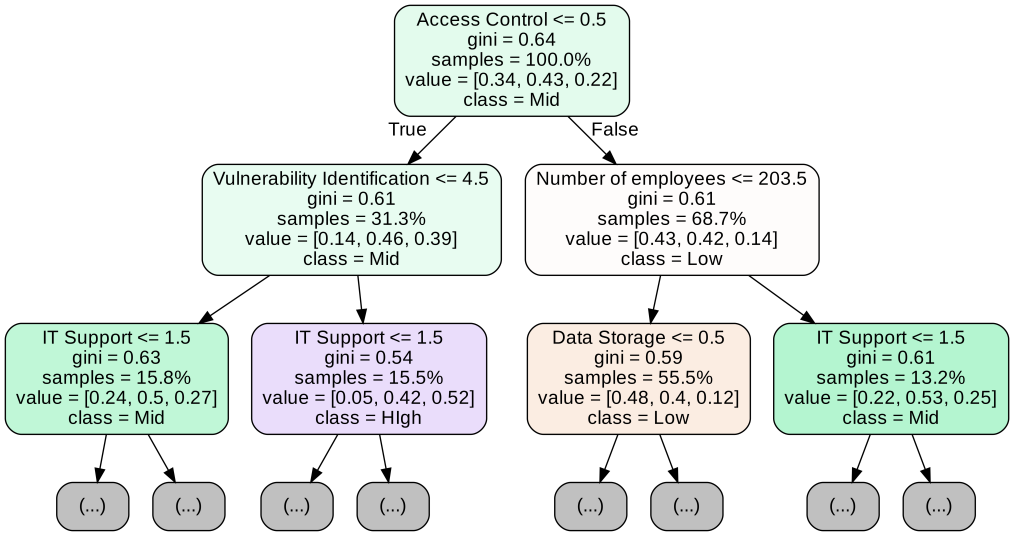


Figure 4.5: Visualization of top three Layers of one of the trees in RF training

Unlike the RF algorithm, there is no specific targeted method to obtain the feature explanation score of NN and SVM. For SVM, this could be achieved with a linear kernel. This is however more difficultly attainable when using an RBF kernel, as the data is implicitly mapped into a higher-dimensional space, and then separated by some hyperplane. The link between the higher-dimension points and the original features cannot be retrieved. It would also be feasible to output the distance between the test instance and the decision boundary, and it can be seen as a confidence score since the probabilities of the class cannot be computed by SVM. However, this score cannot be directly converted into an estimation of the class probability. Another way to visualize the classification of SVM could be by using PCA (Principal Component Analysis) to extract the most important features. However, again, no clear decision boundary for the data points in three-dimensional space was found for the model.

As no targeted method was found to be satisfying, the permutation importance is used to get more insight into both NN and SVM. Permutation importance works as follows: the original model is evaluated on the given data. Then, each column is randomly shuffled one after the other, and the model is re-evaluated. The difference in accuracy between the base model and the one that has been permutated gives the importance of the future of that column [62].

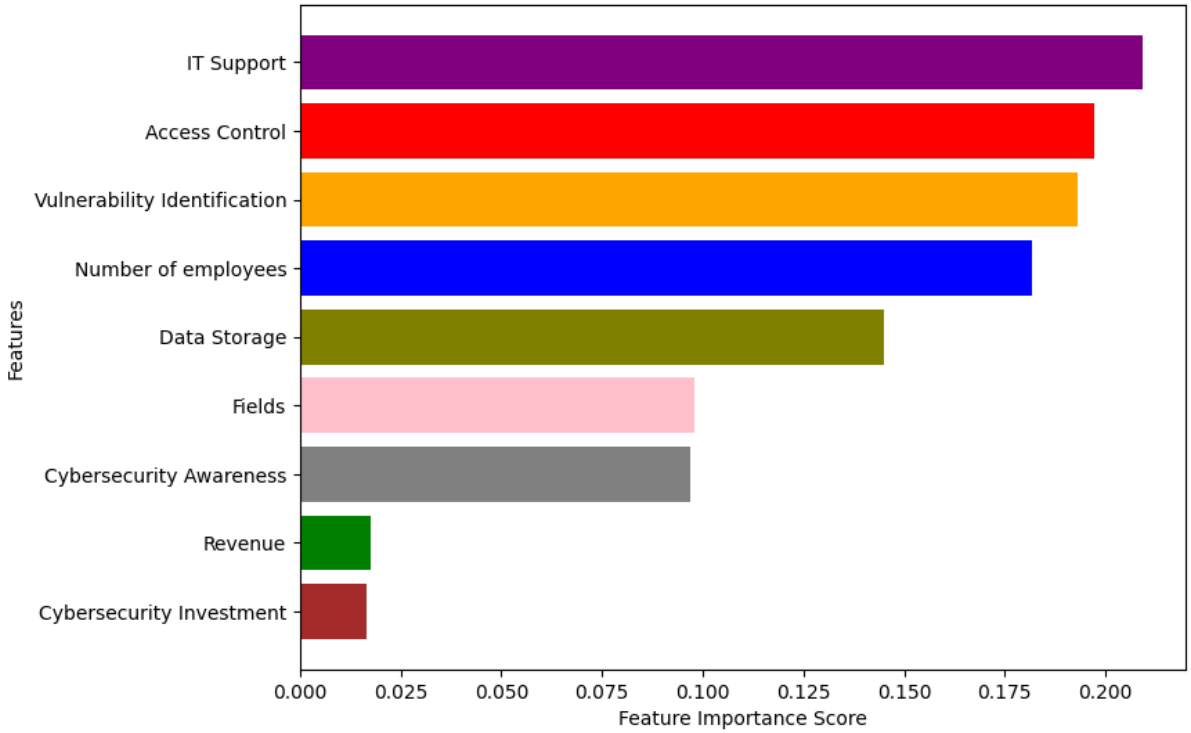


Figure 4.6: The feature importance score generated for NN

As shown in Figure 4.6, the attributes do not have exactly the same order of importance for the NN algorithm as for the RF algorithm, which can be a factor of explanation for the difference in the performance of both models. However, the same trends can be observed: The number of Employees, the most important variable for the RF, is ranked fourth for the NN. Vulnerability Identification is in the top three most important attributes for both, which is again coherent with this attribute having the most impact on the accuracy when left out. Equivalently, Cybersecurity Awareness is in the bottom three of both rankings and thus does not have a high impact on the accuracy when removed.

When comparing the two previous feature importance classifications with the one for the SVM algorithm (Figure 4.7), it is noticeable that the SVM is closer to the NN than to the RF. They have the same order in feature importance, with only slightly different scores. Vulnerability Identification can therefore again be found in the first three most important features, and the Cybersecurity Awareness in ranked among the lower importance attributes. The similarities between the two models could come from the fact that a similar type of feature evaluation was conducted on both of them, but such close results could have been expected seeing the similarities in their accuracy.

In real-world decision making systems, classification models must not only be accurate but also should indicate when they are likely to be incorrect.The NN model, which provides us with the probabilities of each label occurring, can be evaluated from the perspective of the confidence of the model. In other words, a network provides a calibrated confidence measure in addition to its prediction [61].One popular notion of miscalibration is the difference in expectation between confidence and accuracy, i.e. the Expected Calibration Error (ECE):

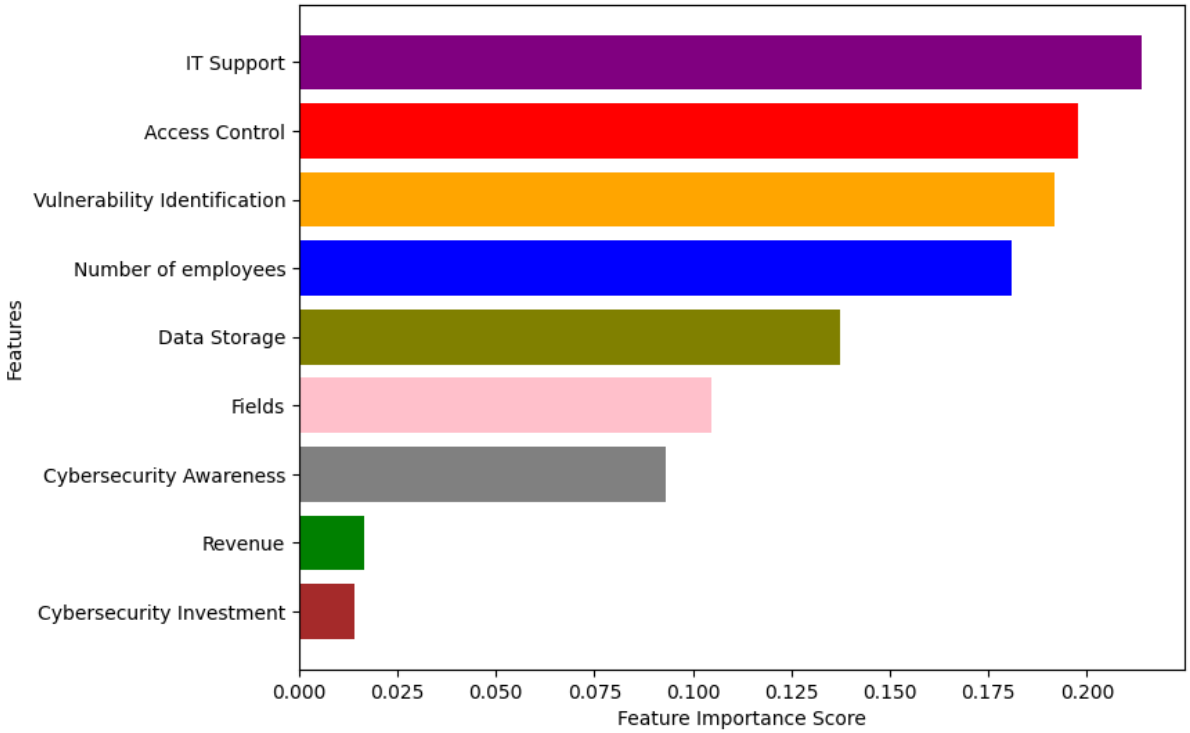


Figure 4.7: The feature importance score generated for SVM

$$\mathop{\mathbb{E}}_{\hat{p}}[|\,\mathbb{P}(\hat{Y} = Y|\hat{P} = P) - p|] \tag{4.3}$$

where X and label Y are random variables, $\hat{Y}$ is a class prediction and $\hat{p}$ is its associated confidence. In this work, the ECE score is approximated by partitioning predictions into 10 equally-spaced bins.

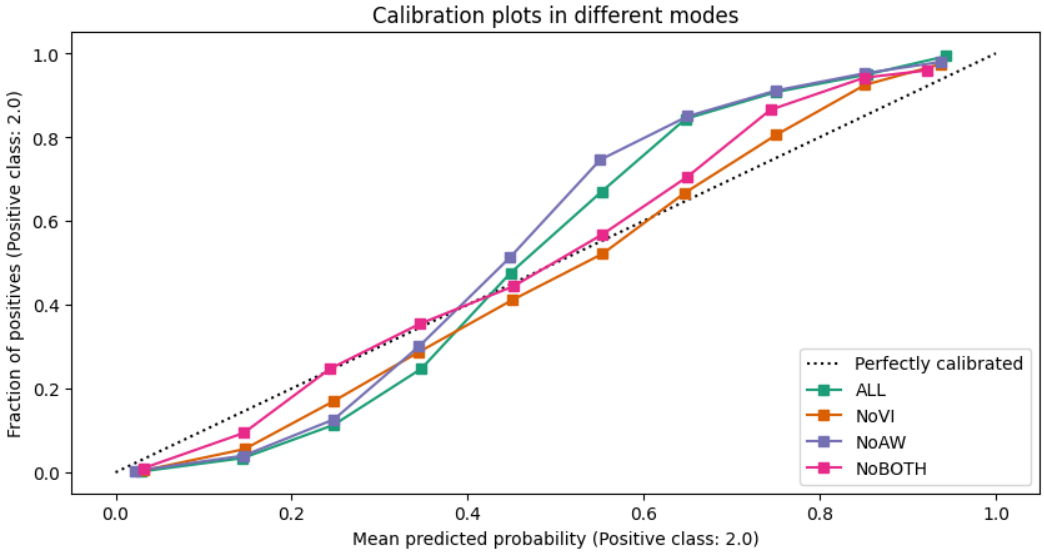


Figure 4.8: The ECE score of NN (with 0.05 std noise)

As shown in Figure 4.8, the diagonal is the perfect calibration line. When it is below the diagonal, the model has overconfidence, i.e. the predicted probabilities are too large. Above the diagonal means that the model is underconfident. As can be seen on the graph, the predicted score is adequately calibrated as the dots fall close to the diagonal line. It is worth pointing out that although the "NoVI" and "NoBOTH" modes perform worse in accuracy and F1 score measurements, their model confidence is more precisely calibrated than the other two models.

4.3 Discussion and Key Findings

General evaluation metrics were applied to the result of the models, as well as some more model-specific methods. The evaluation allows the first period to decide which Gaussian noise to set for the training data, and the second period to assess the accuracy of the models. There were no empirical findings that provided clear evidence of the degree of uncertainty that real-life data should have. Therefore, different options were compared to decide on the final noise. Analyzing the distribution of the accuracy over the data among the three ML models and the equation put into light that a standard deviation of 0.05 was a good compromise, keeping a reasonable accuracy while still having some impact. This decision was further consolidated by the F1 score, as the quasi-linear impact of the noise has a drop point at the 0.05 value.

Comparing the four different modes with the absence of attributes showed a clear difference between the modes All and NoAW, and the modes NoVI and NoBoth. If the Cybersecurity Awareness attribute can be easily ignored, it is not the case for the Vulnerability Identification attribute. The feature importance scores, generated both for the RF and the NN algorithm, provide some explainability as to why there is such a different impact. Vulnerability Identification is respectively the second and third most important feature for RF and NN. Both models rely heavily on that feature to assess the risk, and therefore its absence is detrimental to the final result. Cybersecurity Awareness, on the other hand, is in the bottom three ranking for both algorithms. One can assume that as the performance of SVM is always between the two others, it follows a similar behavior.

Among the three different models, the NN is the best-performing one. This was expected from the related work. SVM is a close second, and RF is largely outperformed. A less foreseeable finding is that NN outperforms the labeling equation constantly even without the addition of any noise, with the exception of the one already present in the attribute definition. This shows that, although the NN is presumably learning the equation to some extent seeing how close both results are, it still manages to outperform the labeling equation by achieving better accuracy.

Although the equation is unsurprisingly good at labeling correctly data according to its own rules, it becomes powerless as soon as attributes are removed. The equation reaches an accuracy that is barely over a random prediction accuracy, between 40 and 50%. This showcases the flexibility that all the models have and the relevance of their use in risk classification.

Finally, it is noticeable that the choice of adding noise directly in the attributes of the data impacts greatly the accuracy. Even without adding any more noise to the equation, the accuracy is capped at around 90%. These results may seem unsatisfactory, especially in comparison with SecRIskAI [4]. However, CyberAlert chooses a different approach than other similar works. The additional noise attempts to bring the synthetic dataset closer to what real world data would be like. The models trained with such data are therefore more robust than if they were trained with overly good data. Thus, what CyberAlert seemingly loses in precision is compensated by being closer to existing data and therefore thrives to be able to maintain this range of accuracy when confronted with it.

Chapter 5

Summary and Conclusions

This project designs and implements CyberAlert, an ML-based cybersecurity risk assessment tool for enterprises to predict their cybersecurity risk. The whole project comprises three parts. First, the synthetic dataset is generated as preparation for ML model training. The value range and correlation between attributes are considered from theoretical and practical perspectives to guarantee the data validity. The appropriate data noise is added in features and labeling.

The second part is ML model selection. TTT chose three supervised learning methods, i.e. RF, SVM and NN for training and predicting cybersecurity risk levels. Based on the synthetic data generated, the model parameters are found and validated with the training and testing dataset. The model optimized parameters are selected by several evaluation measures which include the accuracy and F1-score. Also, model-oriented measures are implemented according to different properties of the model process.

The Ml models serve as the classifiers of the CyberAlert. This prototype asks agents to input the information i.e. the attributes of the companies and provides the option for the agents to skip the fuzzy attributes. Besides that, manipulation of the ML models is allowed so that agents are able to change the configurations of the models within a reasonable parameter range.

One main aspect of future work is synthetic data generation. Though many different adaptations and tests have been implemented for the labeling noise selection, the heteroscedastic noise, the noise in the attributes, is left to be verified. Additionally, it is worth pointing out that despite all the noises that were added to the data, three models are only simulating the equation but not over-functioning. This is a defect that cannot be overcome for now since there is no reliable and accessible real-world data. Last but not least, it is found that the model training time grows exponentially when requiring a better performance. Thus in the real-case study, the balance between model performance and training time is rather important and is supposed to take into consideration.

Bibliography

[1] R. R. Rantala: Cybercrime against Businesses; Special Report, Bureau of Justice Statistics, September 2008, Available at https://static.prisonpolicy.org/scans/bjs/cb05.pdf.

[2] Va. Arlington: Gartner Survey Reveals 82% of Company Leaders Plan to Allow Employees to Work Remotely Some of the Time, Gartner, July 2020, Available at https://www.gartner.com/en/newsroom/press-releases/2020-07-14-gartner-survey-reveals-82-percent-of-company-leaders-plan-to-allow-employees-to-work-remotely-some-of-the-time

[3] M. Miller: FBI sees spike in cyber crime reports during coronavirus pandemic, The Hill, April 2020, Available at `https://thehill.com/policy/cybersecurity/493198-fbi-sees-spike-in-cyber-crime-reports-during-coronavirus-pandemic/`

[4] M. F. Franco, E. Sula, A. Huertas, E. J. Scheid, L. Z. Granville, B. Stiller: SecRiskAI: a Machine Learning-based Approach for Cybersecurity Risk Prediction in Businesses; 24th IEEE International Conference on Business Informatics, Amsterdam, Netherlands, June 2022, pp. 1-10

[5] Erion Sula: SecRiskAI: A Machine Learning-based Tool for Cybersecurity Risk Assessment; Universität Zürich, Communication Systems Group, Department of Informatics, Zürich, Switzerland, August 2021.

[6] E. Amir, S. Levi, T. Livne: Do Firms Underreport Information on Cyber-Attacks? Evidence from Capital Markets; Review of Accounting Studies, Vol. 23, Springer, June 2018, pp. 1177-1206.

[7] Allianz (2021). Allianz Risk Barometer - Identifying the major business risks for 2021

[8] Specops: Survey reveals why SMEs are decreasing their investment in cyber security, November 2020, Available at `https://specopssoft.com/blog/survey-why-smes-decreasing-investment-in-cyber-security/`

[9] Positive technoligies (2019), Cybersecurity Threatscape 2019, Available at `https://www.ptsecurity.com/upload/corporate/ww-en/analytics/cybersecurity-threatscape-2019-eng.pdf`

[10] Positive technologies (2018), Cybersecurity threatscape 2018: Trends and forecasts, Available at `https://www.ptsecurity.com/upload/corporate/ww-en/analytics/Cybersecurity-threatscape-2018-eng.pdf`

[11] Koch, Robert (2017), On the Future of Cybersecurity.

[12] SoSafe: Human Risk Review 2022 : An analysis of the European cyberthreat landscape, Available at https://lp.sosafe.de/hubfs/SoSafe\%20-\%20Human\%20Risk\%20Review\%202022\%20-\%20EN-1.1.pdf?__hstc=106398849.5d40dd624380dc6a0b9375b8714139c4.1656945913770.1656945913770.1656945913770.1&__hssc=106398849.1.1656945913770&__hsfp=2852805679

[13] J. Hegde, B. Rokseth: Applications of machine learning methods for engineering risk assessment A review, September 2019

[14] Y. Ma, M. Chowdhury, A. Sadek and M. Jeihani, "Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration (VII) With Artificial Intelligence (AI)," in IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 4, pp. 615-627, Dec. 2009, doi: 10.1109/TITS.2009.2026673.

[15] Özdemir, A.T., Barshan, B., 2014. Detecting falls with wearable sensors using machine learning techniques. Sensors (Switzerland) 14, 10691-10708.

[16] Elnaggar, R., Chakrabarty, K. Machine Learning for Hardware Security: Opportunities and Risks. J Electron Test 34, 183-201 (2018)

[17] Kim, I.-H.; Bong, J.-H.; Park, J.; Park, S. Prediction of Driverâs Intention of Lane Change by Augmenting Sensor Information Using Machine Learning Techniques. Sensors 2017, 17, 1350

[18] Christopher, A. Balamurugan, Suganya. (2014). Prediction of warning level in aircraft accidents using data mining techniques. Aeronautical Journal. 118. 935-952. 10.1017/S0001924000009623

[19] NIST. NIST Special Publication 800-30 : Guide for Conducting Risk Assessments, Septembre 2012 https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf

[20] Christopher Alberts, C., Dorofee, A., Stevens, J., Woody, C., Introduction to the OCTAVE ® Approach, August 2003, Available at https://apps.dtic.mil/sti/pdfs/ADA634134.pdf

[21] Harris, S. Regulatory Requirements and Risk. Pearson IT Certification, June 2010, Available at https://www.pearsonitcertification.com/articles/article.aspx?p=1594876

[22] Violino, B. 5 IT risk assessment frameworks compared. CSO, , November 2021, Available athttps://www.csoonline.com/article/2125140/it-risk-assessment-frameworks-real-world-experience.html

[23] Craigen, D., Diakun-Thibault, N., Purse, R. 2014. Defining Cybersecurity. Technology Innovation Management Review, 4(10): 13-21.

[24] Cybersecurity. Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/cybersecurity

[25] Material of the Lecture "Foundation of Data Science", Chapter 16 : Neural Networks by Pr. Dan Olteanu, University of Zürich, Fall 2021

[26] Material of the lecture "Introduction to Reinforcement Learning" by Pr. Eleni Vasilaki, University of Zürich, Spring 2022

[27] S. Wang: Artificial Neural Network. Interdisciplinary Computing in Java Programming, Springer International Series in Engineering and Computer Science, Vol. 743, 2003, Boston, MA.

[28] M. Franco, B. Rodrigues, C. Killer, E. J. Scheid, A. De Carli, A. Gassmann, D. Schoenbaechler, B. Stiller: WeTrace: a Privacy-preserving Tracing Approach; KICKS, IEEE ComSoc, Journal of Communications and Networks, Vol. 1, No. 1, October 2021, ISSN 1976-5541, pp. 1-16.

[29] M. Franco, B. Rodrigues, E. J. Scheid, A. Jacobs, C. Killer, L. Granville, B. Stiller: SecBot: a Business-Driven Conversational Agent for Cybersecurity Planning and Management; 16th International Conference on Network and Service Management (CNSM 2020), Izmir, Turkey, November 2020, pp. 1-7.

[30] J. von der Assen, M. F. Franco, C. Killer, E. J. Scheid, B. Stiller: CoReTM: An Approach Enabling Cross-Functional Collaborative Threat Modeling; IEEE International Conference on Cyber Security and Resilience, Rhodes, Greece, July 2022, pp. 1-8.

[31] Muriel Franco: CyberTEA: a Technical and Economic Approach for Cybersecurity Planning and Investment; Universität Zürich, Communication Systems Group, Department of Informatics, Zürich, Switzerland, February 2023.

[32] Gupta, B. B.; Tewari, Aakanksha; Jain, Ankit Kumar; Agrawal, Dharma P. (2016). Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications, (), -. doi:10.1007/s00521-016-2275-y

[33] B. Rodrigues, M. Franco, G. Parangi, B. Stiller: SEConomy: A Framework for the Economic Assessment of Cybersecurity; 16th Conference on the Economics of Grids, Clouds, Systems, and Services (GECON 2019), Leeds, UK, September 2019, pp 1-13.

[34] M. Franco, E. Sula, B. Rodrigues, E. Scheid, B. Stiller: ProtectDDoS: A Platform for Trustworthy Offering and Recommendation of Protections; International Conference on Economics of Grids, Clouds, Software and Services (GECON 2020), Izola, Slovenia, September 2020, pp 1-12.

[35] M. Franco, J. Von der Assen, L. Boillat, C. Killer, B. Rodrigues, E. J. Scheid, L. Granville, B. Stiller: SecGrid: A Visual System for the Analysis and ML-Based Classification of Cyberattack Traffic; IEEE 46th Conference on Local Computer Networks (LCN 2021), Edmonton, Canada, Virtually, October 2021, pp 1-8.

[36] "Beneath the Surface of a Cyberattack: Deloitte: Risk Services." Deloitte, March 11, 2021. Available at https://www2.deloitte.com/global/en/pages/risk/cyber-strategic-risk/articles/beneath-the-surface-of-a-cyberattack.html.

[37] Baeldung. "Multiclass Classification Using Support Vector Machines." Baeldung on Computer Science, August 25, 2021. Available at https://www.baeldung.com/cs/svm-multiclass-classification.

[38] M. Rosenthal: Must-Know Phishing Statistics: Updated 2022, Tessian, January 2022, Available at https://www.tessian.com/blog/phishing-statistics-2020/

[39] A. Unni: How Much Should You Invest In Cybersecurity?, StickmanCyber, January 2018, Available at https://www.stickmancyber.com/cybersecurity-blog/how-much-should-you-invest-in-cybersecurity

[40] IBM Security X-Force Threat Intelligence Index 2022, IBM, Available at https://www.ibm.com/downloads/cas/ADLMYLAZ

[41] Spear Phishing vs. Phishing, Whaling, and Cloning, CoFense, Available at https://cofense.com/project/phishing-vs-spear-phishing

[42] A.C. Singh, K.P. Somase, K.G. Tambre: Phishing: A Computer Security Threat; International Journal of Advance Research in Computer Science and Management Studies, Vol.1, Issue 7, December 2013, pp.64-71

[43] Olivo, C.K., Santin, A.O., Oliveira, L. (2013). Obtaining the threat model for e-mail phishing. Appl. Soft Comput., 13, 4841-4848.

[44] Cybersecurity for SMEs - Challenges and Recommendations, ENISA, June 2021, Available at https://www.enisa.europa.eu/publications/enisa-report-cybersecurity-for-smes

[45] Pros and Cons of In-House Cloud Management vs. Outsourcing, Vault Networks, August 2017, Available at https://www.vaultnetworks.com/pros-and-cons-of-in-house-cloud-management-vs-outsourcing/

[46] Cloud vs. Local File Storage, CommunityIT, Available at https://communityit.com/cloud-vs-local-file-storage-which-is-more-secure/

[47] Vulnerabilities on the corporate network perimeter: Results of automated security assessment, Positive Technologies, October 2020, Available at https://www.ptsecurity.com/upload/corporate/ww-en/analytics/vulnerabilities-corporate-networks-2020-eng.pdf

[48] R. Izquierdo: Do You Need an IT Department at Your Small Business?; The Ascent, August 2022, Available at https://www.fool.com/the-ascent/small-business/it-management/articles/it-department/

[49] Bessy-Roland, Y, Boumezoued, A, Hillairet, C (2020). Multivariate Hawkes process for cyber insurance.

[50] Stanescu, G, Danila, A, Horga, M (2018). Econometric Model Necessary For Analysis Of Existing Correlations Between Human Resources And The Financial Performance Of The Enterprises

[51] Ma, Y., Chowdhury, M., Sadek, A., Jeihani, M. (2009). Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration (VII) With Artificial Intelligence (AI). In IEEE Transactions on Intelligent Transportation Systems (Vol. 10, Issue 4, pp. 615-627). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/tits.2009.2026673

[52] Castro, Y., Kim, Y. J. (2015). Data mining on road safety: factor assessment on vehicle accidents using classification models. In International Journal of Crashworthiness (Vol. 21, Issue 2, pp. 104-111). Informa UK Limited. https://doi.org/10.1080/13588265.2015.1122278

[53] Li, Z., Kolmanovsky, I., Atkins, E., Lu, J., Filev, D. P., Michelini, J. (2016). Road Risk Modeling and Cloud-Aided Safety-Based Route Planning. In IEEE Transactions on Cybernetics (Vol. 46, Issue 11, pp. 2473-2483). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/tcyb.2015.2478698

[54] Pereira, F. C., Rodrigues, F., Ben-Akiva, M. (2013). Text analysis in incident duration prediction. In Transportation Research Part C: Emerging Technologies (Vol. 37, pp. 177-192). Elsevier BV. https://doi.org/10.1016/j.trc.2013.10.002

[55] Wang, K., Simandl, J. K., Porter, M. D., Graettinger, A. J., Smith, R. K. (2016). How the choice of safety performance function affects the identification of important crash prediction variables. In Accident Analysis amp; Prevention (Vol. 88, pp. 1-8). Elsevier BV. https://doi.org/10.1016/j.aap.2015.12.005

[56] Ãzdemir, A., Barshan, B. (2014). Detecting Falls with Wearable Sensors Using Machine Learning Techniques. In Sensors (Vol. 14, Issue 6, pp. 10691-10708). MDPI AG. https://doi.org/10.3390/s140610691

[57] Ayo-Imoru, R. M., Cilliers, A. C. (2018). Continuous machine learning for abnormality identification to aid condition-based maintenance in nuclear power plant. In Annals of Nuclear Energy (Vol. 118, pp. 61-70). Elsevier BV. https://doi.org/10.1016/j.anucene.2018.04.002

[58] Elnaggar, R., Chakrabarty, K. (2018). Machine Learning for Hardware Security: Opportunities and Risks. In Journal of Electronic Testing (Vol. 34, Issue 2, pp. 183-201). Springer Science and Business Media LLC. https://doi.org/10.1007/s10836-018-5726-9

[59] Franco, M., Rodrigues, B., Killer, C., Scheid, E., De Carli, A., Gassmann, A., Schoenbaechler, D., Stiller, B. (2021). WeTrace: a Privacy-preserving Tracing Approach. Journal of Communications and Networks, 1(1), 1-16.

[60] Mandelli, D., Maljovec, D., Alfonsi, A., Parisi, C., Talbot, P., Cogliati, J., Smith, C., Rabiti, C. (2018). Mining data in a dynamic PRA framework. In Progress in Nuclear Energy (Vol. 108, pp. 99-110). Elsevier BV. https://doi.org/10.1016/j.pnucene.2018.05.004

[61] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17). JMLR.org, 1321-1330.

[62] C. Molnar (2022). In Interpretable machine learning: A guide for making Black Box models explainable

[63] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion, 76:243-297, 2021.

[64] M. Valdenegro-Toro and D. Mori, "A Deeper Look into Aleatoric and Epistemic Uncertainty Disentanglement," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 2022 pp. 1508-1516.

[65] React - a JavaScript library for building user interfaces. A JavaScript library for building user interfaces. (n.d.). Avaliable at https://reactjs.org/

[66] Getting started. Getting Started ; Axios Docs. (n.d.). Avaliable at https://axios-http.com/docs/intro

[67] The React Component Library You always wanted. MUI. (n.d.). Avaliable at https://mui.com/

Abbreviations

ACC	Accuracy
ACR	Access Control Rate
ANNs	Artificial Neural Networks
CIR	Cybersecurity Investment Ratio
CBM	Condition-based Maintenance
DDoS	Phish- ing and Distributed Denial-of-Service
DOM	Document Object Model
DT	Decision Tree
ECE	Expected Calibration Error
EER	Employee Exposure Rate
FBI	Federal Bureau of Investigation
KNN	K-Nearest Neighbour
MCC	Multi-Class Classification
MFA	Multi-factor Authentication
ML	Machine Learning
MUI	Material UI
NIST	National Institute of Standards and Technology
OCTAVE	Operationally Critical Threat, Asset and Vulnerability Evaluation
PCA	Principal Component Analysis
RBF	Radial Basis Function
RF	Random Forest
SMEs	Small and Medium-sized Enterprises
SMS	Short Message Service
SVM	Support Vector Machine
SVR	Support Vector Regression
UI	User Interface

List of Figures

List of Tables