Network Attack Detection for Business Safety

Fadia Abduljabbar Saeed, Ghalia Nassreddine, Joumana Younis

NTU Journal of Engineering and Technology (2024) 3 (1) : 1-11
DOI: https://doi.org/10.56286/ntujet.v3i1.535

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Network attack detection for business safety

Fadia Saeed[1], Ghalia Nassreddine[2], Joumana Younis[3]
1. Technical College Agricultural of Mosul, Northern Technical University, Mosul, Iraq
2. Faculty of Business, Jinan University, Tripoli, Lebanon,
3. CNAM, France

Article Informations

Received: 18-06- 2023,
Revised: 25-09-2023,
Accepted: 07-12-2023,
Published online: 10-03-2024

Corresponding author:
Name: Ghalia Nassreddine
Affiliation : Jinan University of Lebanon
Email: ghalia.nasseredine@jinan.edu.lb

Key Words:
Machine learning,
Networks attack,
Network security,
Data privacy,
Deep learnig,
Support vector machine,
Artificial Neural network

A B S T R A C T

In the technology age, the use of networks has hugely increased. this led to an increment in the number of attackers. A network attack is an try to achieve unauthorized access to personnel of an organization's network, steal data or perform other malicious activity. Machine Learning is a subset of artificial Intelligence techniques that teaches machines to learn from historical information. In this paper, a machine learning-based approach was developed to detect network attacks. Two Machine learning models were used: Support vector machine and Artificial neural network. In this approach, a feature selection step based on the p-value is executed first to reduce the size of the dataset. After that, training and testing steps were performed. The proposed approach was tested on a real dataset collected from Kaggle. Confusion matrix, recall, precision, and f1 score were used to test the performance of the used ML techniques. The result shows the efficiency of this approach.

Introduction

During the past twenty years, the business had a vital transformation due to the enormous development of technology. This transformation is called business digitalization. Digitization is the action of transforming physical records of all information and data on paper into digital copies. Digitalization represents the use of the digital process in the business model instead of analog ones. Therefore, business digitalization can be defined as technology that changes how the organization works across its different sectors with clients and other stakeholders. Indeed, a proper digital transformation process is more prominent than buying a new device or software. It includes the whole organization redefining the business processes [1][2].

Many advantages of business digitalization can be resumed by Lee and Falahat (2019) and Lozic (2019) [3] [4]:

- Improved agility: Businesses experienced in using technology can adjust, alter and capitalize on it.
- Enhance the use of resources: recent and contemporary technologies can help different departments within an organization to communicate and share resources easily.
- Expanded response: New technologies can enhance business and client communication. Therefore, it can assist in predicting potential changing customer needs and growing market conditions.
- Boosts Customer Satisfaction: with business digitization, customers can perform tasks such as accessing, updating or deactivating, and reactivating software without needing to deal with human beings. In addition, clients can receive personalized services that meet their needs.
- Decrease human error: it is the main advantage of going digital. Indeed, by using recent technologies and software, the business can stop errors and decrease the time-consuming of manual data entry and human inefficiencies.

However, this considerable transformation has a main drawback. The expanded use of digital transformation has altered data security. Indeed, cyber-attacks, data breaches, and other cyber events are growing as the threat area grows. Hackers or cyber criminals can throw a cyber-attack using one or more computers against a set of computers or networks. A cyber-attack can hatefully disable computers, steal data, or use a computer as a new tool for other attacks. Therefore, defending against hackers and cyber-attacks has become one of the main concerns of any organization [5] [6].

The massive development of technologies provides businesses with many tools that may help protect from cyber-attacks [7] [8].

Machine learning is one of the recent technologies that has proven its efficiency in many sectors, such as healthcare, education, business, and others. It is a subpart of artificial intelligence defined as the capability of a machine to imitate intelligent human behavior. Artificial intelligence systems can perform complex duties similarly to how humans solve problems [9] [10].

In this paper, the role of machine learning in detecting cyber attach for business safety will be studied. The main objective of this thesis can be resumed by:

a. Describing the digitalization process of business.
b. Defining the main threat to business in the digital age.
c. Define the type of cyberattack.
d. Describe the leading machine learning tools that help in detecting cyber-attacks.
e. Test the efficiency of using a Support Vector Machine (SVM) and Artificial Neural Network (ANN) in detecting network attacks.
f. Study the efficiency of using machine learning in the security field.

This paper is organized as follows. First, the network attacks in business will be presented in Section II. In section III, machine learning will be defined. The proposed approach will be presented in Section IV. The dataset and the results will be described and discussed in Section V. This paper will be concluded in Section VI.

Network attacks in Business

The enormous development and progress of the Internet help many network users. Recently, the internet network has become an essential tool for businesses and individual users. For this reason, building a secure network becomes essential for the company. Network security is associated with computers, networks, software, data, and other digital tools. Building a secure network aims to prevent unauthorized access and modification.

In this section, the role of networks in Businesses will be presented first. Later, the network attacks will be described.

A. Network in Business

Networks offer many opportunities to businesses that may evolve. Below are some opportunities [11]:

Business Contacts

Business contacts mainly consist of:

1. Identify and connect with the leading clients in or outside the company.
2. Build relationships with other businesses.
3. Communicate easily with customers.

Business Events

Business events comprise the followings:

1. Share main events like training, seminars, and conferences.
2. Invite connections from a similar central or local area to apply.
3. Analyze promotional opportunities

Information

Information includes the following:

1. Hold sensitive and personal information on a secure server.
2. Authorized staff can only access this information.
3. Protect the personal information of all staff.

Networking groups

These groups work as follows:

1. Share information quickly with colleagues.
2. Find information easily.
3. Meet people from other businesses.
4. Search for people with the same interest.

B. Network Attacks

Network attacks represent any unauthorized action on the digital devices in an organizational network. Usually, hostile parties perform network attacks to change, kill, or steal personal information. Within a network, perpetrators try to get access to internal systems. Network attacks can be classified into two main types [12]:

- Passive network attacks: In this type, hostile access the network and steal information without causing any damage.
- Active network attacks: In this type, hostile modifications damage the data.

An attack can be performed with several tools. These tools will be explained in detail in this section.

Steps of Network Attacks

The steps for performing a network attack are given below (see Figure 1) [12].

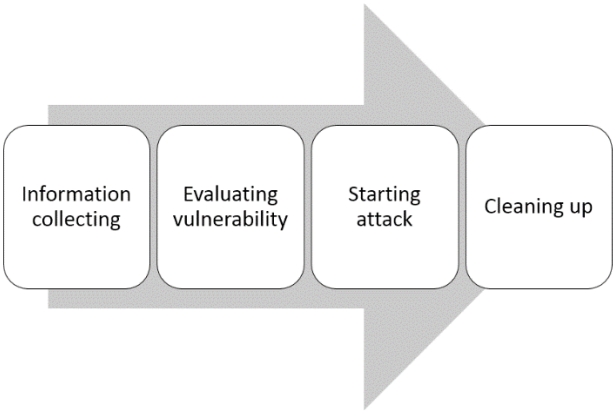


Figure 1: Steps required for a network attack.

- Collecting information: The hostile part should collect information about the network to perform the attack.
- Evaluating vulnerability: Based on the vulnerabilities discovered in step 1, the hostile tries to compromise some nodes in the network by exploiting malicious code as a precursor to the launching of attack(s).
- Starting attack: The attacker throws the attack on the target network using the compromised nodes in step 2.
- Cleaning up: It is the final step of an attack. The hostile should remove all log files from the victim's computers, devices, and network.

Tools of Attacks

- Malware: it is a malicious program like:
  - A Trojan horse (Trojan) **is** a type of malware that hides legitimate content or software**.** When it is within the network, the attacker user can perform all actions as a legitimate user, like opening files, modifying data, and altering the contents of the device
  - Spyware is malware that can install itself on a computer network and catch all online activities without permission. Spyware can privately collects information about network and computer user and send this data to other parties.
  - Ransomware encrypts a device's data and holds it, hostage, for a fee. If the ransom is not paid within a specific time frame, the threat actor threatens to delete or release the valuable data (often opting to sell it on the dark web).
  - In addition to performing its malicious acts, a virus infects other programs and can spread to other systems. A virus is attached to a file and executed when a file is launched. Data and files will then be encrypted, corrupted, deleted, or moved due to the virus. An enterprise-level

antivirus solution can help defend against viruses by protecting the system and data. It is necessary to ensure full scan regularly and update antivirus definitions. They will be activated when a person clicks on a link or attachment. Once the malware is activated, it can [13]:

  - Restrict access to critical network components (ransomware)
  - Install additional malicious software
  - Gather information covertly by transmitting data from the hard drive (spyware)
  - Destroy individual components, rendering the system inoperable

- A distributed denial of service (DDOS) attack is carried out on a website or server to reduce performance deliberately. A large number of computers typically carry out this attack. These computers launch the DoS attack against the website or server. It is finished as follows: Each used computer sends a massive amount of fake demand to the target. The target is inundated with such requests, and the resources become unavailable to respond to valid requests or users [14].
- Phishing uses forged contact information, such as an email, to trick users into opening it and following instructions, such as providing a credit card number. The goal is to steal sensitive data such as credit card numbers and login information or to install malware on the victim's computer [15].
- SQL Injection: A Structured Query Language (SQL) injection is a cyber-attack that involves inserting malicious code into a SQL server. When a server is infected, it releases data. Entering the malicious code into a vulnerable website search box can be simple.
- Password Attacks: A cyber attacker can gain access to a wealth of information with the correct password. Data Insider defines social engineering as "a strategy cyber-attackers use that relies heavily on human interaction and frequently involves tricking people into breaking standard security practices." Accessing a password database or guessing a password are two other types of password attacks.

Consequences of Network Attacks

Network security attacks can lead to many consequences for businesses [16], including:

- Loss of Data: A network attack can lead to the loss of sensitive and essential data, such as employees' financial information or secret contracts with other companies.
- Bad reputation. An attack may damage a company's reputation. It also makes it difficult to regain the trust of customers and other users.
- Decrease or loss of revenue. In many cases, a network security attack can significantly decrease revenue or cause a loss as customers take their business elsewhere.
- Increased costs. Attacks can produce high costs, so a business should hire new staff to update its security systems.

Machine Learning

Machine learning (ML), a subfield of AI, overpowers AI limits. Typically, ML contains all smart systems with performance that may be improved based on. ML systems seek to create an automated analytical model to accomplish cognitive tasks such as classification or image detection. This can be performed by involving algorithms that iteratively improve using a problem-specific training set of data. This set permits computers to uncover hidden patterns without explicitly being programmed. Recently, ML shows high accuracy and precision in many complex problems like regression, clustering, and association problems. Indeed, by learning from prior analyses and removing regularities from massive databases, it will be able to generate reliable and repeatable decisions. Therefore, ML techniques have been successfully used in many sectors, such as weather forecasting, fraud detection, text classification, and others [17]. According to the type of the problem and the availability of data, four main categories of ML can be distinguished (see Figure 2) [18]:

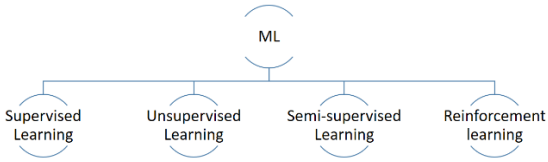


Figure 2: Main categories of ML

Supervised Learning

Supervised learning needs a training dataset that contains inputs and labeled output data. Input and output data of the training set are employed to train the ML model. After that, the training model will be used to predict the output using new inputs [19].

Unsupervised Learning

Unsupervised Learning contains ML systems that can detect patterns without using labeled outputs in the training set [20].

Semi-supervised Learning

Semi-supervised Learning is a general category of machine learning that utilizes labeled data to ground predictions, and unlabeled data to learn the shape of the larger data distribution. Practitioners can achieve strong results with fractions of the labeled data, and as a result, can save valuable time and money [21].

Reinforcement Learning

In this category, the ML system uses the description of the current state of the system and the goal to deliver a list of acceptable actions and their environmental limitations for their outcomes. The ML model executes the process of reaching the goal by employing the regulation of trial and error to maximize compensation.

Table 1 shows some applications of each category.

Table 1: Application of ML categories

ML category	Applications
Supervised Learning	forecast stock markets, Analyze Customer Need, Spam Detection
Unsupervised Learning	Customer segmentation, Basket Analysis
Semi-supervised Learning	Text document classifier, Line finding on GPS data
Reinforcement Learning	A driverless car, Games

A. Machine Learning in Business

ML is used in many business areas such as a bank, clinics, supermarkets, and healthcare providers. Indeed, each business produces data. This data may be analyzed and managed using ML techniques. Below are some examples of using ML in business:

- AI and ML are used in power generation systems to lead to more efficient energy production and prediction utility consumption [22].
- ML techniques could improve production via assembly line automation and can assist in maintaining a safe workplace [23].
- In a telecommunication system, ML could create an outstanding customer experience and saves money for companies [24].
- ML is used also in banking. Indeed, based on natural language processing (technology that lets machines understand human language, verbal or written), conversational AI banking systems can be built. These systems provide customer support [10].
- In the public sector, ML techniques are used to detect tax fraud and also in predicting road traffic accidents [25].

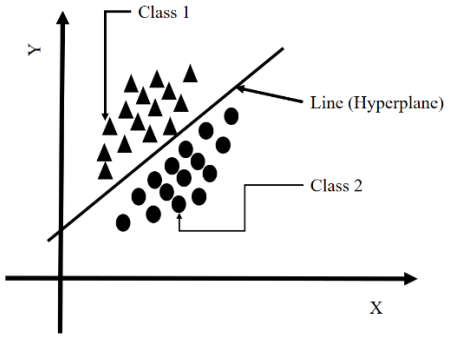
B. Machine Learning Techniques

Many techniques are used in ML applications. In this paper, only two techniques will be presented

Support Vector Machine (SVM)

SVM is a technique of supervised learning that can be applied to classification and regression problems. When it is used in regression problems it is called support vector regression (SVR). The concept of SVM consists of finding a hyperplane in N-dimensional space that separates the different classes of data. The number N is the number of features existing in the dataset [26].

Figure 3 shows an example of SVM in two dimensions (dataset with two characteristics) in which the hyperplane is a single line, having two classes: class 1 represented by triangle points, and class 2 illustrated by circle points. The line correctly separates the two classes. The SVM problem consists of finding the hyperplane with the maximum margin.



X: Independent Variable`
Y: Dependent variable

Figure 3: SVM in two dimensions space

Artificial Neural Network (ANN)

With the huge development of neural networks, many tasks, like image recognition and speech recognition, that were supposed incredibly become convenient now. Indeed, many ML techniques fail precision with a dataset that contains several variables. However, neural networks perform well in this situation. There are two main types of ANN [27].

Single Layer Perceptron (SLP)

SLP is a simple type of ANN. It contains a single layer that connects inputs to outputs via specific weights. It is a feed-forward network (see Figure 4). This type did not perform a backward step.

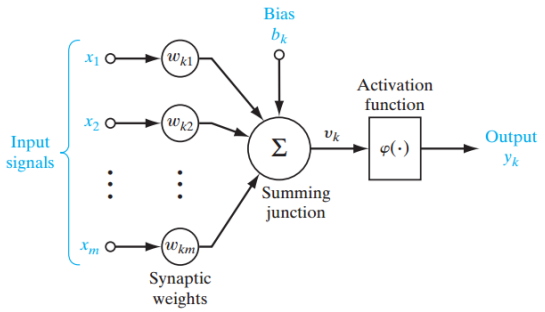


Figure 4: SLP[1]

- (x_j) is the input
- *wki* is the weight associated with inputs. The input with their weight is combined using the following:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j$$

- b_k is the bias. It transforms u_k to v_k using:

$$v_k = u_k + b_k$$

A Multilayer Perceptron (MLP)

MLP may contain one or more hidden layers between the inputs and outputs layers. Each layer is connected to several neurons that interconnect with each other by weight links. The input layers contain neurons equal to the characteristics of the dataset [27]. The output neuron number is equal to the number of classes in the dataset. Figure 5 illustrated an example of MLP with three layers and only one output (class).

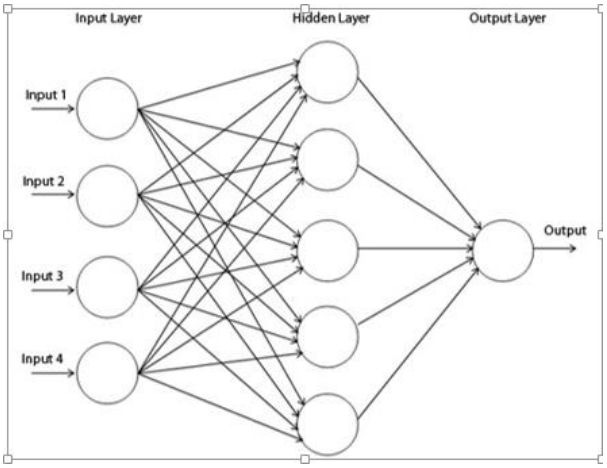


Figure 5: MLP[2]

Figure 5 shows an example of ANN. As illustrated in this figure, three layers exist:

- Input layers that collect input data and send them to the hidden layer
- Hidden layer that manages data and sends the output to the output layer.
- Output layer that contains the result of the ANN.

The initialization of the parameters, weights, and biases is critical in determining the final model. There is a lot of information available about the initialization strategy. A good random initialization strategy will keep you from becoming stuck at local minima. The local minima problem occurs when the network becomes stuck in the error surface and does not stop training even when there is still learning capacity.

Activation function that defines the output of each neuron using:

$$y_k = \varphi(u_k + b_k)$$

where φ(.) is the activation function.

Machine Learning for Network Attacks Detection

The goal of this paper is to detect network using machine learning techniques. Two primary techniques (Support Vector machine, Artificial Neural Networks) are used and compared. Figure 6 illustrates the main steps of detecting network attacks using machine learning techniques:

1 https://www.analyticsvidhya.com/blog/2016/08/evolution-core-concepts-deep-learning-neural-networks/

2 https://www.analyticsvidhya.com/blog/2016/08/evolution-core-concepts-deep-learning-neural-networks/

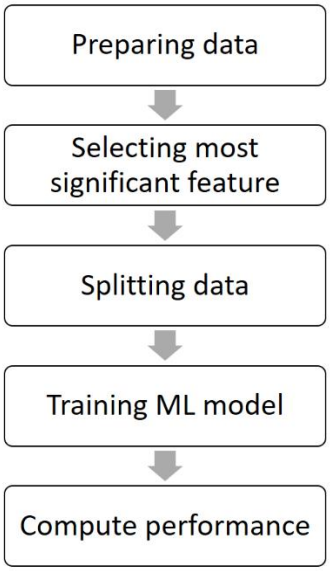


Figure 6: Steps for detecting network attacks using ML

- Preparing data: in this step, the dataset should be cleaned, and removing missing data to be ready to train the ML model
- Selecting the most significant features: usually, the collecting features are not all significant to the type of network attacks. Therefore, removing inappropriate features will be efficient to reduce detection time and increase the detection performance
- Splitting data: to train the ML model, the dataset is divided into:
    - 80% of the data will be used for training the ML model
    - 20% of the data will be used to test the performance of the model
- Compute performance: In this step, confusion matrix and other metrics is used to evaluate the performance of this approach. The confusion matrix (see Figure 7) contains the following data [28]:

		Actual Values	
		Positive	Negative
Predicted Value	Positive	TP	FP
	Negative	FN	TN

Figure 7: Confusion Matrix showing the actual vs. predicted values

- TP (true positive): in this case, the actual and predicted values are positives.
- TN (true negative): in this case, the actual and predicted values are negatives.
- FP (false positive): in this case, the actual value id negative however, the predicted value is positive.
- FP (false negative): in this case, the actual value is positive however, the predicted values is negative.

Using the confusion matrix, the precision, recall, f-measure, and accuracy are computed as shown in Table 2 [28].

Table 2: The formulas adapted to calculate performance metrics

Metric	Description	Formula
Precison	The value of TP elements over TP and FP.	$Precision = \frac{TP}{TP + FP}$
Recall	The value of TP elements over TP and FN	$Recall = \frac{TP}{TP + FN}$
F-measure	The balance of the precision and recall values	$F - measure = 2 \times \frac{Recall \times Precision}{Recall + Precision}$
Accuracy	The ratio of correct predictions to the sample size	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$

Result

In this section, the results of the proposed approach on real dataset will be described. First, the dataset collected from kaggle.com will be presented. After that, the main results will be described and discussed.

A. Dataset

In this paper, a free dataset collected from Kaggle[3] is used. The dataset to be audited was provided which consists of a wide variety of intrusions simulated in a military network environment. It created an environment to acquire raw TCP/IP dump data for a network by simulating a typical US Air Force LAN. The LAN was focused like a real environment and blasted with multiple attacks. A connection is a sequence of TCP packets starting and ending at some time duration between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Also, each connection is labeled as either normal or as an attack with exactly one specific attack type. Each connection record consists of about 100 bytes. For

3 https://www.kaggle.com/datasets/sampadab17/network-intrusion-detection?resource=download

each TCP/IP connection, 41 quantitative and qualitative features are obtained from normal and attack data (3 qualitative and 38 quantitative features). The class variable has two categories: normal vs. anomalous.

B. Most significant features

The most significant feature to select the type of network access will be selected first. This step can be done using the p-value. This value can be computed using the glm function in R:

```
lr<-glm(class~., data = data1, family = "binomial")
summary(lr)
```

The result is illustrated in Figure 10. As shown in this Figure, only 30 features are significant. The other features should be removed from the dataset using:

```
                           Estimate Std. Error z value Pr(>
(Intercept)               -6.210e+00  9.677e-01  -6.417 1.39
duration                  -1.863e-05  2.415e-05  -0.771 0.44
protocol_type             -3.338e+00  3.708e-01  -9.002  < 2
service                    1.730e+01  1.831e+03   0.009 0.99
flag                       1.631e-01  8.832e-02   1.847 0.06
src_bytes                  3.605e-07  1.324e-07   2.722 0.00
dst_bytes                  2.309e-07  1.663e-07   1.388 0.16
land                      -8.458e+00  1.830e+00  -4.621 3.82
wrong_fragment             1.468e+01  8.662e+02   0.017 0.98
urgent                     2.425e+01  2.923e+04   0.001 0.99
hot                        2.171e-01  3.350e-02   6.481 9.12
num_failed_logins          1.688e+00  4.807e-01   3.511 0.00
logged_in                  4.518e-01  2.442e-01   1.850 0.06
num_compromised            1.496e+00  1.201e-01  12.457  < 2
root_shell                 2.445e+00  7.696e-01   3.178 0.00
su_attempted              -4.732e+02  5.393e+02  -0.878 0.38
num_root                  -3.906e-01  1.875e-01  -2.083 0.03
num_file_creations        -2.257e-02  5.989e-02  -0.377 0.70
num_shells                -1.441e+01  5.560e+02  -0.026 0.97
num_access_files          -6.496e-01  7.620e-01  -0.853 0.39
num_outbound_cmds          7.106e-01  3.417e-01   2.080 0.03
is_host_login              0.425e+01  0.923e+04   0.001 0.09
is_guest_login            -5.434e+00  1.171e+00  -4.639 3.50
count                      1.143e-02  1.748e-03   6.539 6.18
srv_count                 -1.174e-02  4.413e-03  -2.661 0.00
serror_rate               -2.185e+00  7.367e-01  -2.965 0.00
srv_serror_rate            7.922e+00  7.639e-01  10.371  < 2
rerror_rate               -2.531e+00  1.033e+00  -2.450 0.01
srv_rerror_rate            7.757e+00  9.236e-01   8.399  < 2
same_srv_rate             -2.751e+00  3.289e-01  -8.365  < 2
diff_srv_rate             -2.026e+00  4.073e-01  -4.974 6.54
srv_diff_host_rate        -1.273e+00  3.469e-01  -3.669 0.00
dst_host_count             1.454e-02  1.045e-03  13.913  < 2
dst_host_srv_count        -2.344e-02  1.221e-03 -19.199  < 2
dst_host_same_srv_rate     3.041e+00  3.386e-01   8.981  < 2
dst_host_diff_srv_rate     7.106e-01  3.417e-01   2.080 0.03
dst_host_same_src_port_rate  3.778e+00  2.305e-01  16.387  < 2
dst_host_srv_diff_host_rate  1.592e+00  5.760e-01   2.765 0.00
dst_host_serror_rate       1.341e+00  3.873e-01   3.463 0.00
dst_host_srv_serror_rate   2.985e+00  4.978e-01   5.996 2.03
dst_host_rerror_rate       1.987e+00  3.342e-01   5.946 2.75
dst_host_srv_rerror_rate  -4.459e+00  5.259e-01  -8.479  < 2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 8: glm() result

After selecting the most siognificant features, the dataset is divided into two main sets:

- Training set (80%)
- Testing set (20%)

In the next sections, the main results of applying SVM and ANN models will be presented.

C. SVM results

Table 4 shows the results of SVM model.

Table 4: Confusion matrix on the testing dataset

		Predicted	
		0	1
Actual	0	2577	21
	1	11	1857

From Table 4, the following points can be concluded:

- The number of normal network access is 2,598
- The number of network attacks is 1,868.
- The number of TP is equal to 1,857. Therefore, SVM can detect 99.411% of network attacks.
- The number of TN is equal to 2,577. Therefore, SVM success in detecting 99.1% of normal network access.
- The number of FP is equal to 21. Therefore, SVM detect 21 access from 2,598 as an attack however, these are normal network access
- The number of FN is equal to 11. Therefore, SVM misses 11 network attacks from 1,868.

For this table, other metrics can be concluded:

- The accuracy of a machine learning model is the metric used to assess which model is best at recognizing relationships and patterns between variables in a dataset based on the input, or training, data. The value of SVM accuracy on training data is equal to 0.971090670170828. Therefore, SVM can detect the correct attack based on input data at the rate of 97.1%.
- Recall, called also true positive rate (TPR), is a value that represents the percentage of data samples that a machine learning model correctly detect to be in the goal class (Positive class) out of the total samples for that class. In this part, the Recall value is equal to 0.988817891373802. Therefore, SVM detects network attacks at the rate of 98.8%. This value is important.
- Precision is a value that represents the accuracy of positive predictions. It is equal to the number of true positive predictions divided by the number of true positive predictions plus false positive predictions. In this case, the value of precision is equal

to 0.9915248967. Therefore, 99.1% of SVM prediction network attacks are correct. It is a good value.

- F-score called also F-measure is a value that combines precision and recall values. In this part, it is equal to 0.990248608.

These results show that the SVM classifier performs well in the network attack problem. In the next section, the ANN performance will be tested on the dataset.

D. ANN results

The resulting confusion matrix of applying ANN model is given in Table 5.

Table 5: Confusion matrix on the testing dataset for ANN

		Predicted	
		0	1
Actual	0	2573	25
	1	17	1851

From Table 5, the following points can be concluded:

- The number of normal network access is 2598
- The number of network attacks is 1868.
- The number of TP is equal to 1851. Therefore, ANN can detect 99.089% of network attacks.
- The number of TN is equal to 2573. Therefore, ANN success in detecting 99.03% of normal network access.
- The number of FP is equal to 25. Therefore, ANN detect 25 access from 2598 as an attack however, these are normal network access
- The number of FN is equal to 17. Therefore, ANN misses 17 network attacks from 1868.

Other metrics can be concluded:

- The accuracy of a machine learning model is the metric used to assess which model is best at recognizing relationships and patterns between variables in a dataset based on the input, or training, data. The value of ANN accuracy on training data is equal to 0.9636902. Therefore, ANN can detect the correct attack based on input data at the rate of 96.36%.
- Recall, called also true positive rate (TPR), is a value that represents the percentage of data samples that a machine learning model correctly detect to be in the goal class (Positive class) out of the total samples for that class. In this part, the Recall value is equal to 0.990365821. Therefore, ANN detects network attacks at the rate of 99.036%. This value is important.
- Precision is a value that represents the accuracy of positive predictions. It is equal to the number of true positive predictions divided by the number of true positive predictions plus false positive predictions. In this case, the value of precision is equal to 0.9908948967. Therefore, 99.08% of ANN prediction Network attacks are correct. It is a good value.
- F-score called also F-measure is a value that combines precision and recall values. In this part, it is equal to 0.9904105068.

E. Comparison between SVM and ANN Performance

Table 6 shows a comparison between ANN and SVM based on recall, precision, accuracy, and F1 score.

Table 6: comparison between ANN and SVM

	Precision	**Recall**	**Accuracy**	**F1 score**
SVM	99.15	98.81	97.19	99.02
ANN	99.08	99.03	96.36	99.04

According to Table 6, the following facts can be concluded:

- ANN has a higher recall than SVM. It denotes the proportion of data samples that a machine learning model correctly identifies as belonging to a class of interest—the "positive class"—out of all samples in that class. As a result, ANN detects genuine network attacks better than SVM.
- The precision of SVM is higher than ANN. Therefore, the quality of a positive prediction made by SVM models is better.
- The accuracy of the SVM model is better. Thus, the fraction of predictions of the SVM model is better.

According to these results, we can conclude that the SVM model is better than the ANN model. The author explains this result as the number of hidden layers is small in the ANN model. However, the author was not able to increase the number of hidden

layers due to the enormous running time of the ANN method. The execution time of these models is illustrated in Table 7:

Table 7: Comparison of running time between ANN and SVM

Method	Run time (second)
SVM	3.5
ANN	151

As illustrated in this table, the execution time of the SVM is smaller than ANN. Therefore, SVM has the best result and the smaller execution time

Conclusion

In the last decade, the use of the internet has significantly increased due to the enormous technological evolution and the quick developments in the internet and communication areas. The resulting increase in threats number has made it difficult for network security to reliably identify breaches. Furthermore, it is impossible to ignore the existence of intruders who intend to conduct a variety of attacks against the network. To effectively detect intrusions across the network, machine learning (ML) and deep learning (DL) was used. ML and DL are a subfield of artificial intelligence science that consists of creating a system that learns from experience. Many ML and DL systems have recently been introduced as good solutions for network attacks.

In this thesis, a new approach of ML and DL for detecting network attacks was proposed. After describing the different types of network attacks and the existing method for protection against these attacks, the author presents the proposed method. It is composed of five main parts: After preparing data, the most significant features were selected based on the p-value. The dataset is split into training and testing sets. Based on the most significant features, the ML model was trained using a training set. After that, a testing set is used to test the performance of the ML model. The proposed approach was tested on real data collected from Kaggle.com. These data are composed of 25,192 rows and 42 features. The p-value test shows that only 30 features were significant. The result shows that:

- The Machine learning model (SVM) performs well in network attack detection with a precision value of 99.15%, recall value of 98.81%, F1 score of 99.02%, and accuracy of 97.19%
- The Deep learning technique (ANN) has a good performance in detecting network attacks with an accuracy of 96.36%, precision of 99.08%, recall of 99.03, and f1 score of 99.04%.
- The SVM is more accurate than ANN. However, the recall value of ANN is better. As a result, the percentage of data samples accurately identified as belonging to a class of interest—the "attack class"—is higher in the ANN model. But the fraction of overall correct predictions of the SVM model is better.

After performing this study, the author suggests the following points:

- Perform more comparison with other ML and DL techniques like Random forest and convolutional neural network.
- Educate more the business staff on the risks of the network attacks
- Perform regular conferences about the types of attacks and how we can avoid them
- Perform training of employees about artificial intelligence techniques and how we can use them in their daily lives.

References

[1] Rachinger, M., Rauter, R., Müller, C., Vorraber, W., & Schirgi, E. (2018). Digitalization and its influence on business model innovation. Journal of Manufacturing Technology Management.

[2] Ritter, T., & Pedersen, C. L. (2020). Digitization capability and the digitalization of business models in business-to-business firms: Past, present, and future. Industrial Marketing Management, 86, 180-190.

[3] Lee, Y. Y., & Falahat, M. (2019). The impact of digitalization and resources on gaining competitive advantage in international markets: Mediating role of marketing, innovation and learning capabilities. *Technology Innovation Management Review*, *9*(11).

[4] Lozic, J. (2019). Core concept of business transformation: from business digitization to business digital transformation. *Economic and Social Development: Book of Proceedings*, 159-167.

[5] Alladi, T., Chamola, V., & Zeadally, S. (2020). Industrial control systems: Cyberattack trends and countermeasures. *Computer Communications, 155*, 1-8.

[6] Huang, K., Siegel, M., & Madnick, S. (2018). Systematically understanding the cyber attack business: A survey. *ACM Computing Surveys (CSUR), 51*(4), 1-36.

[7] Aggarwal, P., Gonzalez, C., & Dutt, V. (2020). HackIt: a real-time simulation tool for studying

real-world cyberattacks in the laboratory. *Handbook of Computer Networks and Cyber Security: Principles and Paradigms*, 949-959.

[8] Gopalakrishnan, T., Ruby, D., Al-Turjman, F., Gupta, D., Pustokhina, I. V., Pustokhin, D. A., & Shankar, K. (2020). Deep learning enabled data offloading with cyber attack detection model in mobile edge computing systems. *IEEE Access*, *8*, 185938-185949.

[9] Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*(2), 183-193.

[10] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29.

[11] Kuzlu, M., & Popescu, O. (2020). Upgrading of a data communication and computer networks course in engineering technology program.

[12] Hoque, N., Bhuyan, M. H., Baishya, R. C., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network attacks: Taxonomy, tools and systems. *Journal of Network and Computer Applications*, *40*, 307-324.

[13] Agrawal, M., Singh, H., Gour, N., & Kumar, M. A. (2014). Evaluation on malware analysis. *International Journal of Computer Science and Information Technologies*, *5*(3), 3381-3383.Networks and Cyber Security (pp. 949-959). Springer, Cham.

[14] Mousavi, S. M., & St-Hilaire, M. (2015, February). Early detection of DDoS attacks against SDN controllers. In *2015 International conference on computing, networking, and communications* (ICNC) (pp. 77-81). IEEE.

[15] Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *Computers & Security*, *68*, 160-196.

[16] Sayeed, S., Marco-Gisbert, H., & Caira, T. (2020). Smart contract: Attacks and protections. *IEEE Access*, *8*, 24416-24427.

[17] Ray, S. (2019, February). A quick review of machine learning algorithms. In *2019 International Conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.

[18] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

[19] Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review*, 1–20.

[20] Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management*, *80*, 52–77.

[21] Ouali, Y., Hudelot, C., & Tami, M. (2020). An overview of deep semi-supervised learning. *arXiv preprint arXiv*:2006.05278.

[22] Akhter, M. N., Mekhilef, S., Mokhlis, H., & Mohamed Shah, N. (2019). Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Generation*, *13*(7), 1009-1023.

[23] Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S. (2019). A review of machine learning for the optimization of production processes. The *International Journal of Advanced Manufacturing Technology*, *104*(5-8), 1889-1902.

[24] Smys, S. (2019). DDOS attack detection in a telecommunication network using machine learning. Journal of Ubiquitous Computing and Communication Technologies (UCCT), 1(01), 33-44.

[25] Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. IEEE Access, 10, 72504-72525.

[26] Tanveer, M., Rajani, T., Rastogi, R., Shao, Y. H., & Ganaie, M. A. (2022). Comprehensive review on twin support vector machines. Annals of Operations Research, 1-46.

[27] Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). Clinical eHealth, 4, 1-11.

[28] Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216-231.