Use of Neural Networks for Predicting Cyberattacks

Bohdan Bebeshko[a], Karyna Khorolska[a], Nataliia Kotenko[a], Oleksander Kharchenko[a], and Tetyana Zhyrova[a]

[a] *Kyiv National University of Trade and Economics, 19 Kioto str., Kyiv, 02000, Ukraine*

Abstract

Cyberattack in the general sense is the use of technical shortcomings of security mechanisms of modern cyberspace to disrupt the work of its elements. From a criminological point of view, a cyberattack should be expressed in the form of an act that involves interfering with the components of information and telecommunications systems and their software or unauthorized modification of computer data through information and telecommunications networks to disrupt the operation of their elements. It takes time to plan defensive actions to prevent the negative impact of cyberattacks. The earlier a future cyberattack is predicted, the more time there will be to plan defense measures. Unfortunately, in the vast majority of cases, data on harmful activities are lacking or incomplete before such activities begin. This article provides research on modern algorithms for cyber-attack prediction working along with data prediction systems for fixing gaps in the incomplete data breach. Afterward, this research was aimed to create a basis for an actual programmatic prototype of a cyber-attack prediction tool. Obtained results demonstrate the model that was investigated can provide a good indication of probable cyber-attack threat based only on its initial behaviors, even in the cases when the given approach has not been exposed to a particular cyber threat before or when it is exposed to an incomplete or damaged dataset for learning purposes.

Keywords

Cybersecurity, cyber-attacks, cyber threads, KNN, SVM, MLP, WFAA, WSAA.

1. Introduction

As you know, it is better to prevent than to eliminate the consequences. This applies not only to everyday life but also to the field of IT. Cybersecurity is the collection of policies, techniques, technologies, and processes that work together to protect the confidentiality, integrity, and availability of computing resources, networks, software programs, and data from attack. Cyber defense mechanisms exist at the application, network, host, and data levels.

The Law of Ukraine "On Basic Principles of Ensuring Cyber Security of Ukraine" of October 5, 2017, provides the following definition of the concept of cyber-attack: hardware, other technical and technological means, and equipment) and aimed at achieving one or a combination of the following objectives: violation of confidentiality, integrity, availability of electronic information resources processed (transmitted, stored) in communication and/or technological systems, obtaining unauthorized access to such resources; violation of security, sustainable, reliable and regular operation of communication and/or technological systems; use of the communication system, its resources and means of electronic communications to carry out cyber-attacks on other objects of cyber defense" [1]. According to Melnyk S. [2] Cyberattack is one of the largest cyber threats of our time. A cyberattack can be seen as an independent phenomenon, and as the quintessence of cyber warfare or terrorist activities in cyberspace. Cyberattack (cyber-attack) in the general sense is the use of technical shortcomings of security mechanisms of modern cyberspace to disrupt the work of its elements. From

Cybersecurity Providing in Information and Telecommunication Systems, January 28, 2021, Kyiv, Ukraine
EMAIL: b.bebeshko@knute.edu.ua (A.1); k.khorolska@knute.edu.ua (A.2); kotenkono@knute.edu.ua (A.3); a.kharchenko@knute.edu.ua (A.4); zhyrova@knute.edu.ua (A.5)
ORCID: 0000-0001-6599-0808 (A.1); 0000-0003-3270-4494 (A.2); 0000-0002-2675-6514 (A.3); 0000-0002-9255-9287 (A.4); 0000-0001-8321-6939 (A.5)


a criminological point of view, a cyberattack should be expressed in the form of an act that involves interfering with the components of information and telecommunications systems and their software or unauthorized modification of computer data through information and telecommunications networks to disrupt the operation of their elements.

It takes time to plan defensive actions to prevent the negative impact of cyberattacks. The earlier a future cyberattack is predicted, the more time there will be to plan defense measures. Unfortunately, in the vast majority of cases, data on harmful activities are lacking or incomplete before such activities begin.

A key problem in cyberattack forecasting with unconventional signals is that not all signals produce values on a regular basis since many are event-driven. This adds to the potential errors due to incorrect sensor readings, unavailability of observation for some time, or problems in the data processing pipeline. An intelligent signal imputation method is needed to deal with signal sources that do not report values for a significant period. In addition, these signals may have different significant lags i.e. time elapsed between the observed public data to the cyber incident. A systematic way to capture the diverse significant lags with imputed signal values is non-trivial and requires a system-level design. Adding more to the challenge, successful cyber incidents are expected to be rare events for a reasonably protected organization, resulting in imbalanced data. Imbalanced data can lead to biased or inaccurate models where the predictive power of unconventional signals is not captured. This paper provides a comprehensive treatment of all these problems individually and as an integrated system. The overall system is tested using the cyber incident data provided by an anonymized company nicknamed K9.[20, 26]

2. The Aim

Due to the high integration of communication and information technologies in the power grid, an extra layer has emerged at every level of the grid – generation, transmission, distribution, consumption – enabling the acquisition, storage, analysis of data through networked sensors, measuring and processing units to improve quality production and delivery of power. This layer determines high volumes of data that utility companies need to manage with appropriate tools. The term that best describes the set of high volumes of data managed only by advanced methods of analysis and knowledge extraction on high power processing units is “Big Data” [22, 25].

This scientific research is aimed to analyze existing algorithms for cyber-attack prediction. Therefore as one speech about prediction using neural networks the problem of the incomplete or damaged data should be taken into consideration as well.

Therefore the main aim of the article was to provide research of modern algorithms for cyber-attack prediction working along with data prediction systems for fixing gaps in the incomplete data breach. Afterward, this research was aimed to create a basis for an actual programmatic prototype of a cyber-attack prediction tool.

3. Models and Methods

As was mentioned, we have used data examples from already available datasets. The data set was obtained via a feature extraction mechanism from a large data provider.

It is a field that investigates how simple models of cyber-attacks prediction can be used to solve difficult computational tasks like the predictive modeling tasks we see in machine learning. The goal is not to create realistic models of the cyber-attack, but instead to develop robust algorithms and data structures that we can use to model difficult problems.

The power of neural networks comes from their ability to learn the representation in training data and how to best relate it to the output variables to be predicted. In this sense, neural networks learn mapping. Mathematically, they are capable of learning any mapping function and have been proven to be a universal approximation algorithm.

Let one overview several most common algorithms.

3.1. Multilayer Perceptron

Multilayer perceptron (MLP) can be designed by connecting the individual perceptron into neural network-based architecture. MLP (Fig. 1) is recognized as a category of Artificial Neural Network because all input and middle layers provide input to their further layers MLP is a class of forwarding artificial neural network. Artificial neural networks are a machine learning method that takes their roots from the idea of the way the human brain works, like learning and obtaining new information, and making decisions based on information gained before. Each MLP includes at least three layers: an input, hidden, and output layer. For training purposes, MLP uses a supervised learning technique called back-propagation.

Multilayer perceptron (MLP) is characterized by the presence of one or more hidden layers, with computation connections called hidden neurons, whose function is to intervene between the external inputs and the network output in a useful manner. To extract high-order statistics, more hidden layers may be added. The network acquires a global perspective despite its local connectivity due to the extrasynaptic connections and the extra dimension of neural network interconnections [23].

Cömert and Kocamaz mentioned that before network training, it is necessary to understand the amount of data because of the size of the data neurons in the neural network. During the network evaluation, 70% of the data set is used for training, and the weights and deviations can be updated according to the network and the target output value; 15% is used for verification so that the network stops training before overfitting occurs; 15% is used as testing to predict the performance of the network [24].

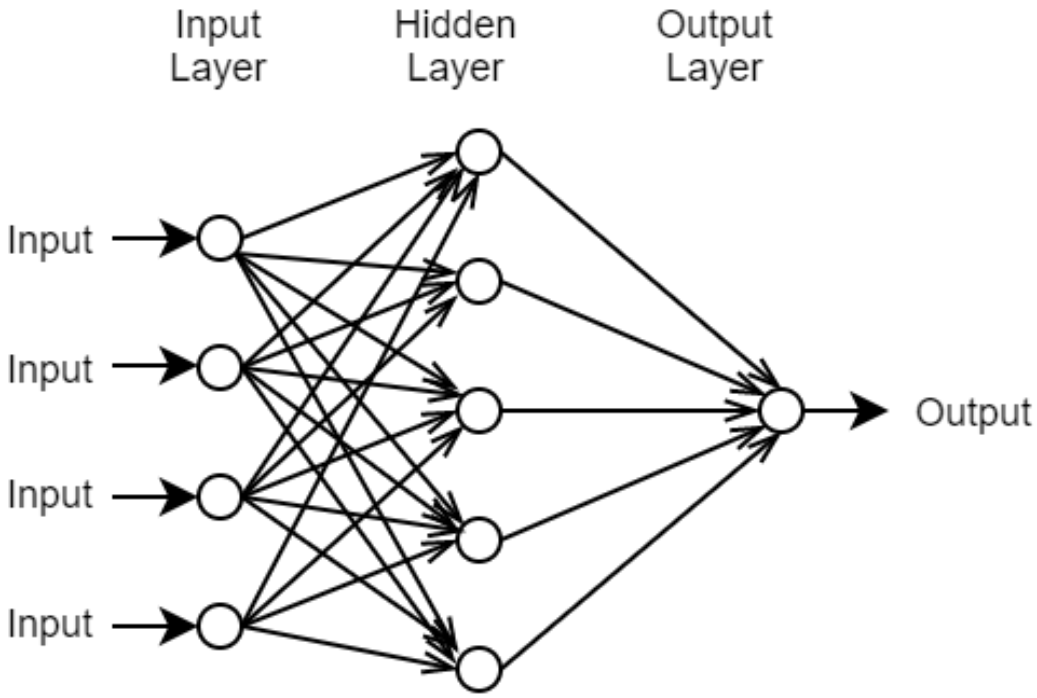

Figure 1: Graphical representation of the MLP principles.

3.2. Support Vector Machines

Basically the cyber-attack detection is a classification problem, in which we classify the normal pattern from the abnormal pattern (attack) of the system. In the machine learning sphere, support vector machines (SVM) are supervised learning models with associated learning algorithms that process and further analyze data flow that is in turn further used for classification and regression metrics. Set of training examples that is given for learning purposes, moreover, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear

classifier (nevertheless there are also methods such as Platt scaling created to use SVM in a probabilistic classification model). In other words SVM is a machine learning algorithm that learns to classify data flow using labelled training data samples getting into one or two classes. The SVM algorithm forms a model that is able to predict whether a new example can get into one category or the other. Fig. 2 represents a hyperplane described by (w, b), where w is a weight and b bias displayed in a finite space of the training node N with according points:

$$\{(x1, y1)(x2, y2 \ \ldots \ (xN, yN))\}$$

Where $xi \in R\hat{}d \ and \ yi \in \{1, -1\}$. This is conducted since in general the larger the margin, the lower the generalization error of the classifier.

One can figure that previously authors within their experiments [11] were utilizing GA for pre-processed KDD Cup 99 dataset in a preprocessing module for data reduction, however, it was heavily complicated to process the dataset with all 40features at a time. Therefore GA was used to select 10 definite features out of all available 40 features represented in the KDD Cup 99 dataset and applied SVM for classification of the resulting dataset. The experiment was carried out with 500,000 samples from the dataset out of which 90% was used as training data for algorithm learning purposes and the remaining 10% as test data. The classification process was running until a 10 fold cross-validation was done for possible results verification. The SVM model classified four different attacks (DoS, probe, U2R, R2L attacks). Another practical experiment [12] proposed a modern method integrating PCA and SVM by optimizing the kernel parameters using an automatic parameter selection approach. The experiment was performed on KDD Cup 99 dataset that contained five categories of digital traffic (normal, DoS attack, R2L attack, U2R attack, and probe attack). Each network record had 40 definitive features of which 7 were discrete and 33 continuous features. C parameter for RBF kernel of SVM was optimized by the proposed automatic parameter reduction along with cross-validation to reduce the training and testing time to give better accuracy in detecting attacks.

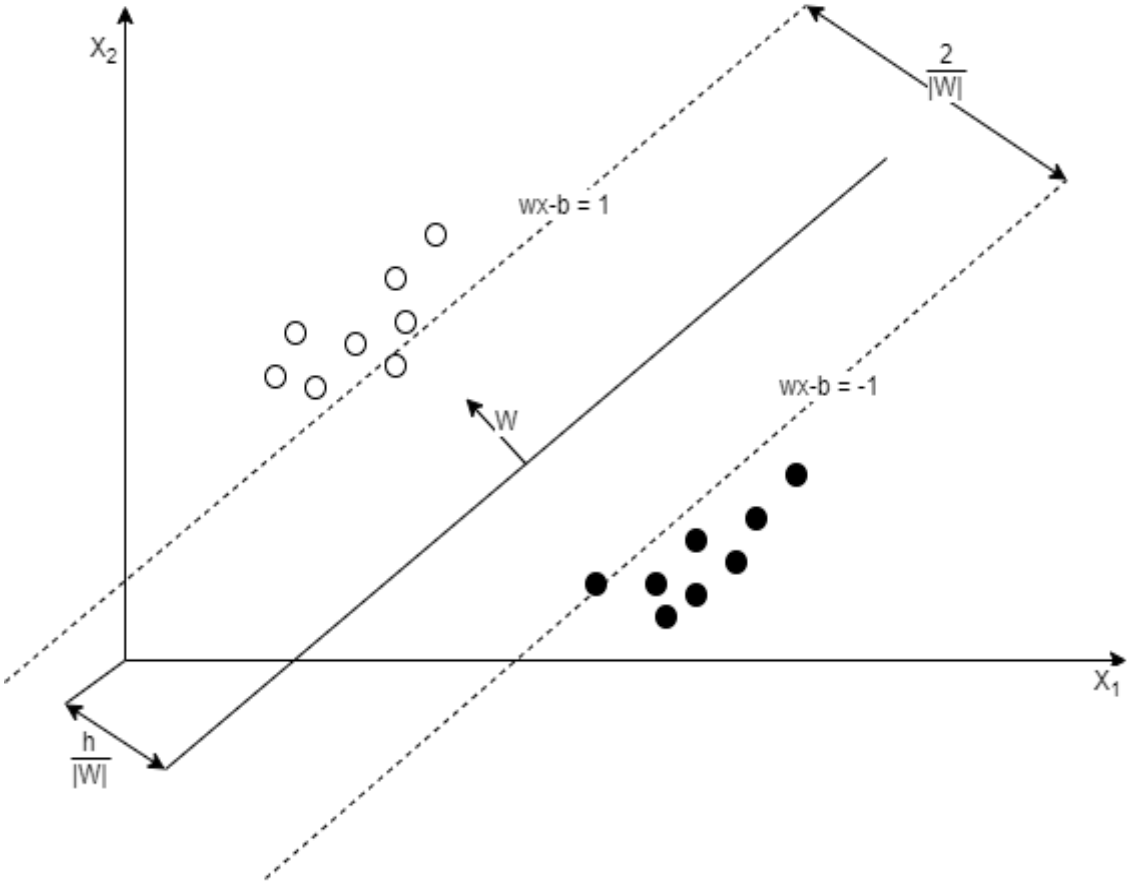

Figure 2: Maximal margin plane for SVM model samples.

So, basically, SVM algorithm can be shown as follows:

SVM Algorithm
Input: Train Data Set—Train, Test Data Set—Test
Output: Cyberthreats classification:

1. Read train data set.
2. Apply SVM algorithm.
3. Generate SVM model for kernel function.
4. Read test data set.
5. For each characteristic in test data.
6. Extract all the features.
7. Apply SVM algorithm.
8. Return result of test data.
9. End.

3.3. K-nearest Neighbors

KNN (Figure 3) is one of the simplest and most effective supervised learning algorithms. It is widely used for searching through the available dataset in order to associate new data points with similar existing points [13]. KNN, which provides satisfying performance over multidimensional data and is a fast algorithm during the training process, is relatively slow in the estimation point. K-nearest neighbors are the nonparametric classification that stores available data and classifies new data based on how similar they are in terms of distance. In the early 1970s, KNN was considered one of the most advanced nonparametric techniques in statistical prediction and estimation as well as pattern recognition [14, 15].

Therefore, this technique is known to be non-parametric and highly efficient in classification [16, 21]. It evaluates the class labels of the test samples [17] based on the majority of test sample neighbors. The parameter k is determined by the user. Based on the test sample, k numbers of training points are determined by taking the closest distance to the test sample. The prediction of the test sample is the k nearest neighbors [18]. A hybrid method of intrusion detection system by the author of [17] proposed a combination of K-NN and GA algorithms. They tested the hybrid algorithm on KDD Cup 99 dataset labeled out of five classes; normal, probe attack, DoS attack, R2L attack, and U2R attack. The dataset was reduced to 7,000, 8,000, 9,000, 10,000, and 15,000 records respectively with 19 features. GA was used to select the k nearest neighbor for K-NN classifier.

Three experiments were performed to conclude on the hybrid method. Each experiment used 10 fold cross-validation with different k values. The values in terms of accuracy were compared with the conventional K-NN. The proposed hybrid method proved better than conventional K-NN in all values of k used in the experiment. On the same dataset author [19] compared the performance of the K-NN and SVM model and RIPPER method in detecting attacks. The multi-attribute decision was adopted in this experiment. In the classification of an unknown document vector X, k –nearest neighbor algorithm ranks the document's neighbor among the training document vectors and uses the class labels of the k most similar neighbors to predict the class of the new document. The similarity in each neighbor to X is used to determine the classes of the neighborhood where the similarity is measured by the Euclidean distance between two document vectors. With this adoption, they categorized each new program behavior in the dataset into either normal or attack classes. Each system call was treated as a word and each process as a document. k were varied between 15 and 35 till an optimal value of 19 was found. The classification was performed with the K-NN and SVM models. The Hit rate was compared to the RIPPER method. The results showed 95.30% accuracy rate and 7.01% false alarm rate for KNN and SVM model whilst the RIPPER method gave 86.30% accuracy rate and 8.2% false-positive rate.

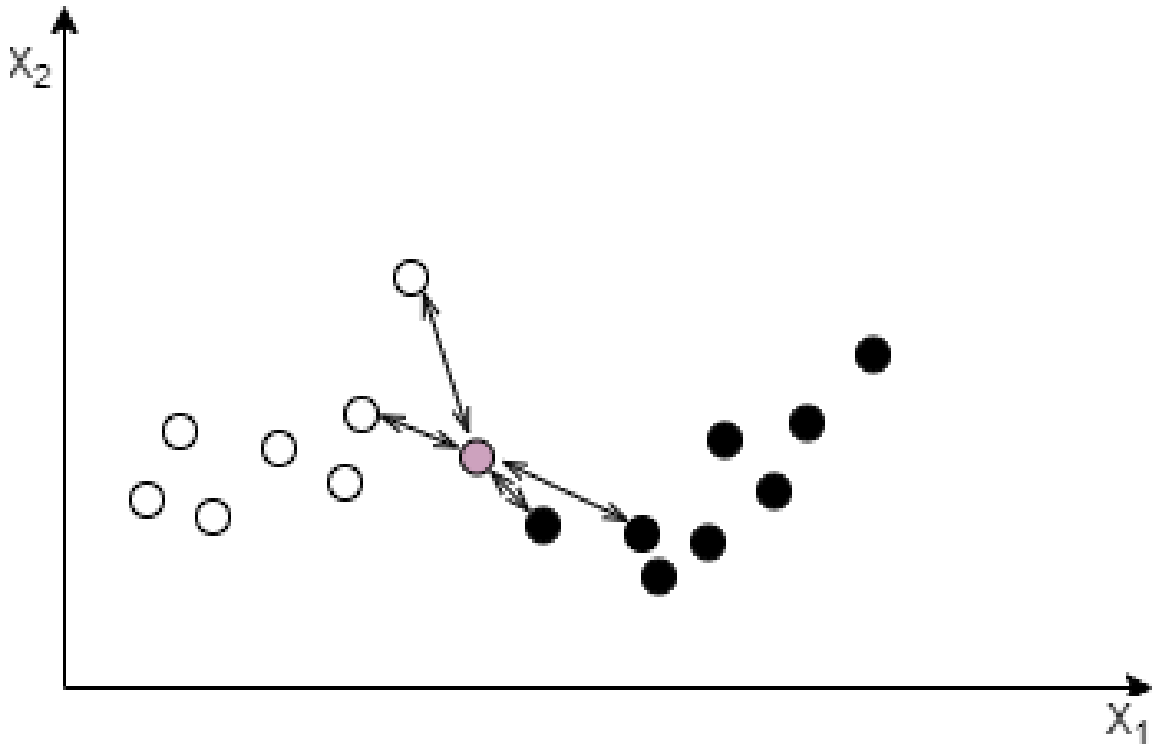

Figure 3: Representation of the KKN algorithm.

4. Experiment Results

Some hosts are designed to detect previous steps in a cycle of attacks that do not involve the actual network infrastructure. Besides, the determination of significant flows can be automated for the respective hosts to enable the prediction system to adapt to different characteristics of non-traditional flows.

It is worth noting that, unlike previous work on predicting cyberattacks, this article emphasizes the importance of incomplete use of data, which is often underestimated when assessing the actual cyberattacks in the organization.

Missing host values in the data set can affect the quality of the learning process and impair the efficiency of detection algorithms. There are different approaches to dealing with incomplete data. One of the simplest solutions is to ignore missing host values; however, in this case, the number of training copies of the forecasting system may decrease, which will lead to poor performance assessments. An alternative method is to replace the missing values with the average value of the existing missing values. Another way to consider is a more sophisticated approach that represents or replaces a missing value with a forecasting strategy.

Saar-Tsechansky and Provost [3] noted that different types of substitution methods may be better than others depending on the circumstances. Rahman and Davis [4] used the average rule induction algorithm, decision trees called J48, KNN (K-Nearest Neighbor), and SVM (Support Vector Machines) methods to replace missing values in the data set and concluded that machine methods training works better on other substitution methods. Luengo [3] used fourteen substitution methods for missing values and found that substitution methods lead to greater performance than approaches that ignore missing values. Supporting previous findings in the literature, they also concluded that there is no one-size-fits-all method of substitution. Farhangfar [4] made a comprehensive review of existing replacement methods and developed a unified structure aimed at encapsulating a set of methods. They divide the methods of processing missing values into three categories, where missing data in the data were lost or missing, the most preferred algorithms were used, and missing values were predicted using average replacement methods or machine learning methods. However, as far as we know, this has been little interpreted in the field of cybersecurity. We will review a set of

incomplete host replacement methods to replace missing values and demonstrate an increase in prediction power for predicting cyberattack incidents.

As mentioned earlier, missing data is one of the biggest problems of machine learning and has a significant impact on the learning and forecasting process. Data may be absent accidentally or systematically when lost value can be observed when a condition is met. In any of the possible cases, there is no data - this is a "traffic jam" and therefore this situation requires a special solution to improve the performance of forecasting techniques. As already mentioned - the simplest solution is to ignore the missing hosts. However, depending on the number of missing hosts, this approach can lead to ignoring a huge set of data and a huge loss in predictive power. The solution is to replace the missing data with different methods. Support Vector Machines (SVM) is one of the training methods that can be used to replace missing data. Another method is the k-nearest neighbor (KNN), which is a classification algorithm also known as the IBC (Instance-Based Classifier). The KNN instance is denoted by the sign of most of its neighbors k, where $k >= 1$. KNN is one of the widely used methods of replacing missing host values. Neural networks can be used in the decision-making process because they are able to model complex nonlinear relationships within a data set. Multilayer Perceptron (MLP) is a neural network direct feedback technique that can analyze the problem of classifying nonlinear functions. The MLP architecture is a complex of at least three layers of nodes and can evaluate data that cannot be linearly separated.

To enable the prediction of a cyberattack with incomplete data, you can use the predictive signal (host) technique (PSI) based on SVM, MLP, and KNN algorithms to replace the lost values in the hosts. However, all errors (failures) need to be aggregated. The input set used in this paper is a set of interval signals whose values change over time. The use of interval signals as predictors leads to the emergence and necessity of processing large amounts of data. To reduce the amount of data, instead of using the entire history of intervals as input data for the forecasting model, it is possible to aggregate a number of anomalous observations and provide only this aggregate value of the forecasting algorithm. Three algorithms were used for data aggregation.

The first algorithm is "Aggregation and aggregation based on runtime" (ALA). This approach to aggregation covers all types of errors that are above the significance threshold $\frac{2}{\sqrt{n}}$ where n is the number of paired measurements. The aggregation period (delta) t is calculated based on the execution time T_1 (1 is one interval list), the last signal that correlates with the main absolute truth (absolutely correct signal) and the smallest error value T_m. The signal is aggregated using the average value of the historical values that fall into the aggregation cycle, so it is possible to represent the expression:

$$X_t = \frac{1}{\Delta_t} \sum_{i=T_l}^{T_m} X_{t-i}, \quad (1)$$

The next algorithm is the Weighted Total Mean Aggregation (WFAA). The weighted total average method takes into account all measured errors and weighs them according to their correlation coefficients. The value of y corresponding to each error is a measurement of the correlation between the lagging signal and the current measurements of true truth. Weighing the signal values allows you to significantly correlate the measurements for aggregation, without ignoring any data in history. This algorithm can be described by an expression:

$$X_t = \frac{1}{w} \sum_{i=1}^{w} X_{t-i} \cdot c_i, \quad (2)$$

where w is the number of measured errors, c_i is the correlation coefficient.

The last algorithm used is the Weighted Significant Mean Aggregation (WSAA). The weighted significant mean approach takes into account only highly correlated errors in calculating the averaged data. As in the previous method, this allows each significant error to affect the aggregation relative to the general correlation with the truth. Two methods for determining the threshold V for critical correlation were considered.

1. Tabular significance (t-WSAA).

This method is used t-tables with $\alpha = 0.5$ for calculation V_a (3).

$$V_a = \sqrt{\frac{qt(\frac{\alpha}{2}, n-2)}{qt(\frac{\alpha}{2}, n-2)^2 + (n-2)}}, \quad (3)$$

where $qt()$ – quantile function, n – the length of the interval. The critical data set of method a is defined as $C = \{(X_i, c_i) | c_i > V_a\}$.

2. Significance of the value (v-WSAA).

This method describes the value—v of each error based on its correlation coefficient C_i to $P_i < 0.5$ (which is considered a significant correlation). $p(r)$ is represented by the expression 4.

$$p(r) = \frac{e^{\frac{-1}{2}\cdot\left(\frac{r^2}{\sqrt{\frac{1-r^2}{n-2}}}\right)}}{\sqrt{\frac{1-r^2}{n-2}} \cdot \sqrt{2\pi}}, \quad (4)$$

where r is the correlation coefficient understudy, n is the length of the interval. The critical data set of the method a is defined as $C = \{(X_i, c_i) \mid p(c_i) < 0.5\}$

The significant average is calculated using the signal values and correlation coefficients contained in the critical data set, thus obtaining equality 5.

$$X_t = \frac{1}{|C|} \sum_{i=1}^{|C|} C(i).X \cdot C(i).c, \quad (5)$$

The proposed algorithm accepts the following input data: allInstances - a set of incomplete data, the percentage of most cases for deletion – p, the number of nearest neighbors that must be taken into account to obtain k – as input and output data of a new data set with an even distribution of majority and minority.

- Function algo(allInstances, p, k)

Let majInstances – be a set of the majority of copies in all objects of Insistence;
Let minInstances – be a collection of minority cases at all intervals;
Let sMin – розмір minInstances;
Let sMaj – розмір majInstances;

- Finding the first minority cluster using K-Means clustering using Euclidean metrics.

```
K := 2;
minorityClusterFound := false;
while minorityClusterFound != true do
let clusters – be the first set of K clusters in allInstances;
if clusters include a minority cluster then
let cMin – be the centroid of a minority cluster in clusters;
minorityClusterFound := true;
else
K := K + 1;
end
if K == sMin then
break;
end
if  minorityClusterFound != true then
let cMin – be the average of all minority cases in minInstances;
```

- Filter most majInstances objects: remove p percent of majInstances closest to cMin;

let majInstancesNew – be a set of majInstances after filtering;

- Redefining the significance of objects majInstances

majWeight := 100/(100 - p);

- Establish the significance of each object majInstancesNew = majWeight;
- Redefining the significance of objects minInstances

minWeight := sMaj / sMin / 2;

- Establish the significance of each object minInstances = minWeight;
- Generation of new minWeightsMin minors using k-NN

let minInstancesSyn – be a set of created artificial specimens of minorities
return majInstancesNew ∪ minInstances ∪ minInstancesSyn;

The predicted power of different host replacement methods is compared using the ALPHA data set for EM (Endpoint Malware), MD (Malicious Destination), and ME (Malicious Email) attack types. 58% of the cases in the data set contain at least one lost host, and substitution methods such as SVM, KNN, and MLP are used to replace these lost host values. After replacing the lost host values, the Bayes.Net classifier with 10-fold cross-validation is used to calculate the AUC (Under Receiver Curve) value for each type of attack.

To enable cyberattack forecast with incomplete data, a novel predictive signal imputation technique (PSI) that is based on the SVM, MLP, and KNN algorithms is used to fill in the missing values in the signals. It was shown that KNN performs better than other approaches and the proposed predictive imputation method helps to improve the prediction performance of a BayesNet classifier in terms of the AUC.

The significance of the unconventional signals may not always be the same. To consider the significant observations more than the insignificant ones, across correlation-based signal aggregation approach (ASL) is used to aggregate signals over the past significant lags. Several approaches including ALA, WFAA, WSAA-t, and WSAA-p are compared using the K9 data set and it was shown that WSAA-t helps to improve the cyberattack prediction performance in terms of the AUC.[20]

The obtained AUC values are shown in Table 1. The values in the "None" row show the AUC values when the replacement method was not used, which means that the instances with lost hosts are obsolete. You may notice that replacement methods increase prediction power compared to a situation where instances with lost host values are deleted. Also, the k-Nearest Neighbor algorithm works much better than SVM and MLP for cyber-attacks EM, MD, and ME.

Table 1
AUC values for different PSI approaches

	EM	MD	ME
SVM	0.60	0.64	0.78
KNN	0.88	0.91	0.95
MLP	0.73	0.55	0.83
None	0.51	0.46	0.74

According to the test of each method, different PSI methods are applied to the ALPHA data set under the same conditions. Replacing missing signals has been shown to increase model performance by up to 87%, 90%, and 96% AUC for predicting endpoint attacks, malware, and malicious e-mail, respectively. The results show the reliability of cyberattack prediction, where the integrated results provide approximately $0.6 - 1.0$ F-measurements over time. The proposed structure allows to the assessment of the compliance of non-traditional signals for predicting cyberattacks. Careful integrated use of PSI without excessive use of replaced signals to determine significant lags can provide even better and more reliable performance.

5. Further Research

Obtained results demonstrate that the model that was investigated can provide a good indication of probable cyber-attack threat based only on its initial behaviors, even in the cases when the given approach has not been exposed to a particular cyber threat before or when it is exposed to an incomplete or damaged dataset for learning purposes. Further researches and development processes will be sharpened on the following key points:

As for now, the approach was tested only using the Windows10 environment. Windows10 is a good operating system, but it is not a production server-side operating system. Therefore the first step of future development will be an optimization of the algorithm to function under the Linux environment. So it will be completely rewritten from C# to more low-level programming languages such as C or GoLang.

The next step will be algorithm optimization in the field of resource efficiency. Since testing was held on a home PC with Ryzen 5 3600X on board with 32 GB of RAM, testing consumes 100% of the CPU and RAM capacity. Therefore it is obvious that algorithms should be heavily optimized.

6. Conclusions

The idea of Big Data and wide Internet use is considered as an opportunity to provide a more reliable and accurate source for business intelligence. However, the versatile characteristics of Big Data and the huge need for the use of open Internet connections possess the potential to compromise the reliability and integrity of data.

Cybersecurity is considered one of the serious challenges for researchers. Therefore, in this study, we have proposed a more reliable and accurate ensemble-based approach to classify benign and malicious activities to identify and prevent possible cyber threats. Our proposed approach is highly accurate and able to classify (between benign versus malicious) with an accuracy of 0.993. In the future, this study will be further investigated to identify the threat pattern in cybersecurity

This paper was aimed to detect network attacks by using machine learning methods. The assessment of the existing approaches was done. And the combination of most profitable was figured out and represented in the previous section of the article. Further research plans were declared in order to clarify the importance of the proper system composition before it can actually be named as a production-ready system.

7. References

[1] On the basic principles of cybersecurity of Ukraine, 2017. URL: https://zakon.rada.gov.ua/laws/show/2163-19/ed20171005#n10

[2] S.V. Melnyk, O. Tikhomirov, On the problem of forming the conceptual and terminological apparatus of cybersecurity, in: Actual problems of information security management of the state, Kyiv, Department of the Security Service of Ukraine, 2011, pp. 43–48.

[3] M. Saar-Tsechansky, F. Provost, Handling missing values when applying classification models, Journal of machine learning research 8 (2007) 1623–1657.

[4] M. M. Rahman, D. N. Davis, Machine Learning-Based Missing Value Imputation Method for Clinical Datasets, in: Yang, G.-C., Ao, S., Gelman, L. (Eds.), IAENG Transactions on Engineering Technologies: Special Volume of the World Congress on Engineering, Springer Netherlands, Dordrecht, 2012. pp. 245–257. doi:10.1007/978-94-007-6190-2_19.

[5] J. Luengo, S. García, F. Herrera, On the choice of the best imputation methods for missing values considering three groups of classification methods, Knowledge and Information Systems 32 (2012) 77–108. doi:10.1007/s10115-011-0424-2.

[6] A. Farhangfar, L. A. Kurgan, W. Pedrycz, A novel framework for imputation of missing values in databases, IEEE Trans Syst Man Cybern, 37 (2007) 692–709.

[7] V. Lakhno, et al., Clustering network attack features in information security analysis tasks, Electronic Professional Scientific Edition, Cybersecurity: Education, Science, Technique 1 (2020) 45–58. doi:10.28925/2663-4023.2020.9.4558.
[8] L. Tereikovska, Model of formation of study examples of the neural network intended for the analysis of the keyboard handwriting, Electronic Professional Scientific Edition, Cybersecurity: Education, Science, Technique 1 (2020) 104–114. doi: 10.28925/2663-4023.2020.9.104114.
[9] A. Shevchenko, G. Zastelo, E. Shpachynsky, Analysis of application of machine learning methods on the basis of artificial neural networks for detection of cyber threats, Information Technology and Security 7 (2019) 79–90. URL: http://its.iszzi.kpi.ua/article/view/184327/184136.
[10] R. A. Demidov, P. D. Zegzhda, M. O. Kalinin, Threat Analysis of Cyber Security in Wireless Adhoc Networks Using Hybrid Neural Network Model, Aut. Control Comp. Sci 52 (2018) 971–976. doi:10.3103/S0146411618080084.
[11] B. Senthilnayaki, K. Venkatalakshmi, A. Kannan, Intrusion detection using optimal genetic feature selection and SVM based classifier, in: 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), 2015, pp. 1–4.
[12] L. Teng, S. Teng, F. Tang, H. Zhu, W. Zhang, D. Liu, L. Liang, A Collaborative and Adaptive Intrusion Detection Based on SVMs and Decision Trees, in: IEEE International Conference on Data Mining Workshop, 2014, pp. 898–905.
[13] Z. A. Baig, S. Sanguanpong, S. N. Firdous, T. G. Nguyen, C. So-In, Averaged dependence estimators for dos attack detection in iot networks, Future Generation Computer Systems 102 (2019), pp. 198–209.
[14] K. Q. Weinberger, J. Blitzer, L. K. Saul, Distance metric learning for large margin nearest neighbor classification, in: Advances in neural information processing systems, 2006, pp. 1473–1480.
[15] L. I. Kuncheva, Combining pattern classifiers: methods and algorithms: John Wiley & Sons, 2004.
[16] K. Shi, L. Li, H. Liu, J. He, N. Zhang, W. Song, An improved KNN text classification algorithm based on density, in: IEEE International Conference on Cloud Computing and Intelligence Systems, 2011, pp. 113–117.
[17] Y. Canbay, S. Sagiroglu, A Hybrid Method for Intrusion Detection, in: 14th International Conference on Machine Learning and Applications (ICMLA), 2015, pp. 156–161.
[18] H. Zhang, G. Chen, The Research of Face Recognition Based on PCA and K-Nearest Neighbor, in: 2012 Symposium on Photonics and Optoelectronics, 2012, pp. 1–4. doi:10.1109/SOPO.2012.6270975.
[19] Q. Zeng, S. Wu, Anomaly Detection Based on Multi-Attribute Decision, in: WRI Global Congress on Intelligent Systems, 2009, pp. 394–398.
[20] A. Okutan, G. Werner, S. Yang, K. McConky, Forecasting cyberattacks with incomplete, imbalanced, and insignificant data, *Cybersecur* 1 (2018). doi:10.1186/s42400-018-0016-5.
[21] A. Kalizhanova, S. Akhmetov, V. Lakhno, W. Wojcik, G. Nabiyeva, Optimization model of adaptive decision taking support system for distributed systems cyber security facilities placement, International Journal of Electronics and Telecommunications 66 (2020), pp. 493–498.
[22] W. Sun, P. Bocchini, B. D. Davison, Applications of artificial intelligence for disaster management, Nat Hazards 103 (2020) 2631–2689 (2020). doi:10.1007/s11069-020-04124-3
[23] Y. N. Chi, J. Chi, Saltwater anglers toward marine environmental threats using multilayer perceptron neural network framework, Int. J. Data Sci. Adv. , 2 (2020) 6–17.
[24] Z. Cömert, A. F. Kocamaz, A study of artificial neural network training alorithms classification of cardiotocography signals, Bitlis Eren Univ. J. Sci. Technol. 7 (2017) 7 93–103.
[25] V. A. Lakhno, Algorithms for Forming a Knowledge Base for Decision Support Systems in Cybersecurity Tasks, Advances in Intelligent Systems and Computing 938 (2020) 268–278.
[26] V. A. Lakhno, V. G. Malikov, D.Y. Kasatkin, A. I. Blozva, V. G. Saiko, V. N. Domrachev, Computer-Based Support for Searching Rational Strategies for Investors in Case of Insufficient Information on the Condition of the Counterparty, Advances in Intelligent Systems and Computing 1225 (2020) 120–130.