ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Think Like a Machine: Examine and Predict How Cyber Attacks Will Be Detected

[1]Mr. M Rafath Kumar, [2]Margani Rup Tej Kumar, [3]Vegiraju Sunil Varma, [4]P Sai Srinivas Murali Karthik, [5]N Tarun Durga Prasad,

[1]Assistant Professor, Department of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry, E. G. Dist. A.P 533107.

[2,3,4,5]Student, Department of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry, E. G. Dist. A.P 533107

Abstract:

Every day, governments and their populations suffer enormous financial losses due to cybercrime, making it one of the most pressing global challenges. There needs to be a better understanding of the tactics used by cybercriminals as the frequency of these attacks has been rising. Recent developments have brought security models and prediction tools based on artificial intelligence to address the problems of detecting and preventing cyber threats. Despite the abundance of research on crime prediction tactics, it is possible that they may benefit from being better equipped to anticipate cybercrime and cyberattack approaches. A possible approach to this issue may be to use real-world data to pinpoint when an assault happened and who was behind it. All aspects of the crime, the perpetrator's characteristics, the extent of the damage, and the entry points for the assault are covered by this data. Information gathered from victims of cyberattacks may be accessed by forensic teams via application procedures. Using two models and machine learning, this study examines cybercrime and makes predictions about how qualities might help identify the cyber-attack method and the perpetrator. In terms of accuracy, this research found no significant difference between eight distinct machine-learning methods. When compared to other cyber-attack approaches, the Support Vector Machine (SVM) linear model proved to be the most accurate. The initial model provided helpful information about the kinds of assaults that victims should expect to encounter. As far as detecting bad actors goes, logistic regression is the way to go because of its high success rate. In order to forecast identification, the second model compared perpetrator and victim characteristics. Our research shows that as one's education and economic level rise, the probability of falling prey to cyber assaults declines. Because it would make cyber-attack detection easier and the fight against them more efficient, this suggested idea is highly anticipated by cybercrime departments. Subjects: SVM, cybercrime, machine learning algorithms, cyberattacks.

Introduction:

The goal of machine learning is to use historical data to generate predictions about the future. The field of artificial intelligence known as machine learning (ML) allows computers to "learn" new tasks and knowledge without human intervention. Creating computer programs that can adapt to new data is the ultimate aim of machine learning. It is essential to use specific algorithms in the training and prediction procedures. An algorithm is fed the training data and then utilizes that knowledge to make predictions about fresh test data. Three distinct types of machine learning exist. Learning may be categorized into three main groups: supervised, unsupervised, and reinforced. Programs need human tagging of data before they may utilize it for supervised learning. Labels are not used in unlabeled learning. The learning algorithm was made possible by it. The input data clustering must be determined using this approach. As a last point, reinforcement learning evolves via interacting dynamically with its surroundings and absorbing both positive and negative feedback. Using Python, data scientists may use machine learning techniques to discover patterns that provide valuable insights. We may classify these algorithms as either supervised or unsupervised learning, depending on how they "learn" from the data in order to generate predictions. "Classification" is the process that assigns a label to a set of data points.

Classes go by a few different names: goals, labels, and groupings. In the realm of classification, predictive modeling entails creating a near-perfect mapping function between continuous input variables (X) and discrete output variables (y). Classification is a kind of supervised learning in statistics and machine learning; it involves teaching a computer to classify incoming data points according to established rules. Binary information (such as the subject's gender or the email's spam status) or other types of data could be included in this data set. Classification problems manifest themselves in many forms; some examples are file sorting, speech recognition, biometric identification, and handwriting analysis. 2. The second step is to contact the police department's division that specializes in the kind of crime that the public has reported. All of these numbers are recorded in the database of the unit. For each crime, the police meticulously document the details, such as the kind, manner, year, etc. They collect information, sort it into categories, run analysis, and visualize the results. Multiple simultaneous cyberattacks on the same target are only counted as a single incident. Pay more attention to the specifics of the event than to the statistics if you want to see whether alternative approaches were used. While there are several recorded violations, cybercrime has gained significant attention as of late. Little has been done to stop cybercrime, despite the fact that it has caused immense material and moral damage. The lack of empirical research on this topic in previous cybercrime investigations led to its selection as a focus. By analyzing the victim's profile, the suggested model hopes to predict how likely it is that the victim would become a victim of crime. Law enforcement will also be able to better profile cybercrime suspects, victims, and offenders. Unwanted impacts may also be mitigated with the help of the model. In addition to increasing public knowledge of possible hazards, the study's findings will allow for better targeted therapies. Our data collection included actual occurrences of cybercrime in the Elaz province from 2018 to 2022. Acquiring clean data and preparing it for analysis using machine learning algorithms was a difficult task. The data collecting process included investigating all facets of cybercrime. Data science techniques were used to extract the unnecessary parts. Figure 1 displays the data set's overall crime count, damage total, assault count, and attack vector count. In addition, the information about these four characteristics is organized based on color. Using this data from many modules, Python was able to make predictions.

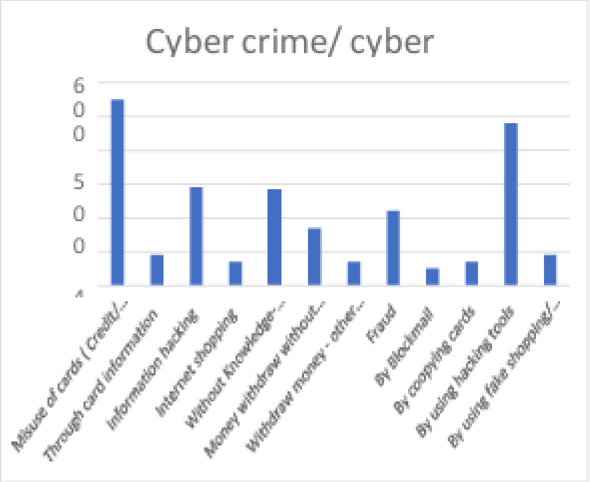


Figure 1: The number of cybercrimes or cyberattacks done by various methods in the datasets.

To display data, this software made use of NumPy, Pandas, and Matplotlib, which is one of the main libraries of the program. According to the article, the key advantages of utilizing machine learning techniques include the following: the ability to spot evolving criminal strategies, the capacity to extract complicated data relationships, the potential to generate outcomes that are beyond human prediction, and the recognition of numerous patterns in both structured and unstructured data. Picking relevant and related characteristics from a dataset is called feature selection. When training machine learning, it helps to save space and time. Poor characteristic selection may lengthen training timeframes, which in turn can raise the model's error rate and make interpretation more difficult. We have defined the features and elements of our dataset. Items pertaining to real crimes are included in Table 1. Fig. 2 displays the characteristics of the data used for training as well as the median, maximum, and minimum values for the attributes in our dataset.

Table 1: Analysis on Crime Type

Item	Crime type

Crime	Debit/ Credit Card utilization;
	Information misuse;
	Hacking

Gender type	Male/ Female/ Other
Age	Below 27 years; 28-38 years age 39-51 years age factor
Income factor	Low/ Moderate/High
Job	Working / self -employ/ House wife/ retired person/ etc.
Marriage	Single / Couple
Education Qualification	Primary/ Highe r/ Under graduate
Harming	Internet shopping without having proper knowledge. Money withdraws by unknown person
Attack	Misuse of ATM card/ Creditcard Social media accounting Digital data hacking Threatening mails Shopping in unauthori zed website

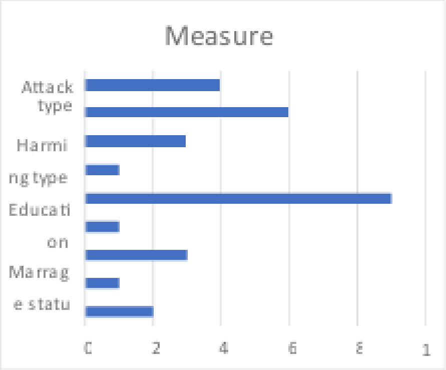
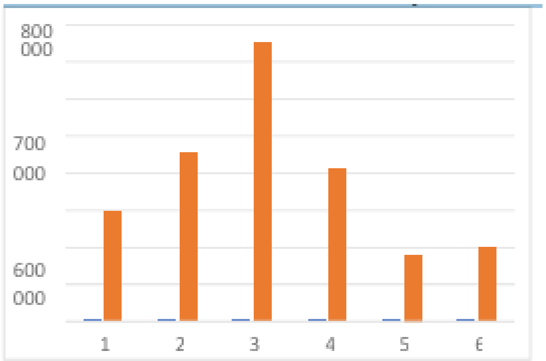


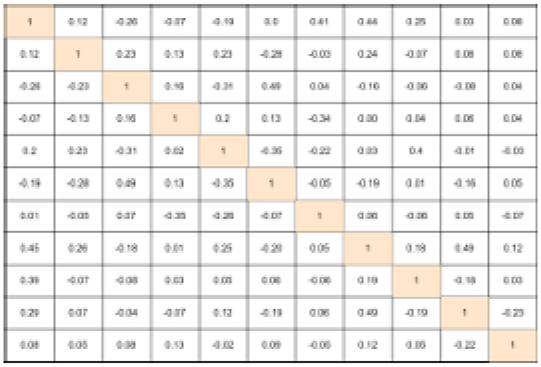
Figure 2: Measurement of actual crime

When features are standardized, they are scaled so that they fit a normal distribution. It is recommended to do this before using any machine learning algorithms. Each column was given a value between 1 and 10, and the data was normalized to represent the diversity of data that was given. To maximize damage, aggressiveness, and attack methods, Python's Standard Scaler() was used. Testing made use of 20% of the data, whereas training made use of 80%. Methodology for Male and Female offenders We were able to anticipate the assault tactic in the first model by integrating details about the crime, the offender, the victim (gender, age, employment, income, marital status, degree of education), and the actual assault. We sought to identify the offender in the second model by looking at their age, gender, income, profession, marital status, education level, assault kind, injury severity, and attack method, among other things. 3. Discussion and Findings: The study's analysis of incident data is effective in reducing criminal activity and apprehending its perpetrators. The purpose of this research is to find ways to reduce criminal activity by analyzing collected data. The findings will clarify the police investigations and expose any hidden details. Machine learning algorithms may analyze victim information, cybercrime tactics, and the status of the perpetrator to identify whether the same cybercriminal was responsible for an attack. The victims of the cyber catastrophe in Elaz province have had their damages calculated using a variety of methodologies. To find the damages for each victim, we added up all the years in the dataset. The decrease in comparable incidents, especially after 2018, is largely attributable to the deterrent effect provided by the rule and awareness campaigns. Cyberattacks in Elaz have resulted in massive financial losses, as seen in Figure 3. It is critical to manage attack strategies and cyber security in light of the losses highlighted above.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025



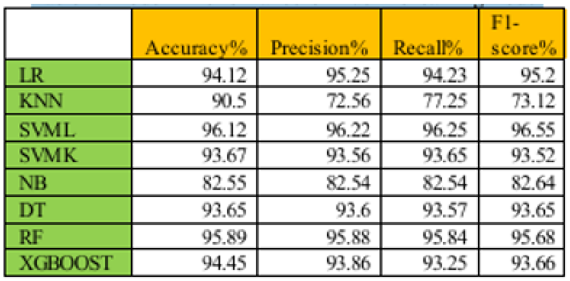
relationships between almost every possible set of variables. Figure 3: Economy damage by cyber attacks Results from SVM (Linear), RF, Logistic Regression, XGBoost, SVM (Kernel), DT, KNN, and NB, among others, are shown here. May determine the Pearson correlation coefficient by referring to the example in Fig. 4. This correlation matrix clearly shows robust relationships between almost every possible set of variables.



1	0.12	-0.26	-0.07	-0.19	0.0	0.41	0.44	0.25	0.03	0.08
0.12	1	0.23	0.13	0.23	-0.28	-0.03	0.24	-0.07	0.08	0.08
-0.26	-0.23	1	0.16	-0.31	0.49	0.04	-0.16	-0.06	-0.09	0.04
-0.07	-0.13	0.16	1	0.2	0.13	-0.34	0.00	0.04	0.06	0.04
0.2	0.23	-0.31	0.02	1	-0.35	-0.22	0.03	0.4	-0.01	-0.03
-0.19	-0.28	0.49	0.13	-0.35	1	-0.05	-0.19	0.01	-0.16	0.05
0.01	-0.03	0.07	-0.35	-0.28	-0.07	1	0.06	-0.06	0.05	-0.07
0.45	0.26	-0.18	0.01	0.25	-0.28	0.05	1	0.18	0.49	0.12
0.39	-0.07	-0.08	0.03	0.05	0.06	-0.06	0.19	1	-0.16	0.03
0.29	0.07	-0.04	-0.07	0.12	-0.19	0.06	0.49	-0.19	1	-0.23
0.08	0.05	0.08	0.13	-0.02	0.09	-0.06	0.12	0.06	-0.22	1

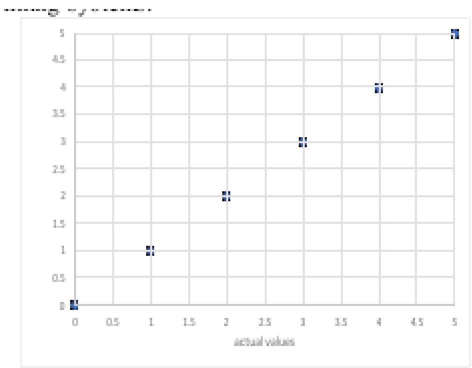
Figure 4: Confusion Matrix

Table 2: Model 1- Performance of Machine learning model

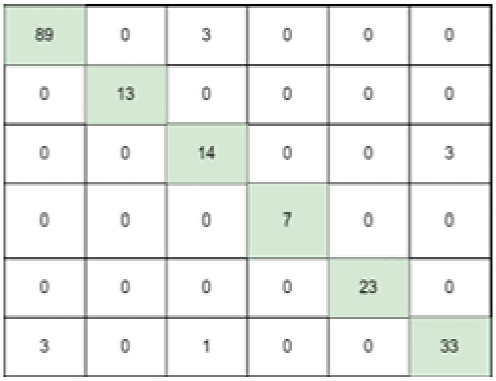


	Accuracy%	Precision%	Recall%	F1-score%
LR	94.12	95.25	94.23	95.2
KNN	90.5	72.56	77.25	73.12
SVML	96.12	96.22	96.25	96.55
SVMK	93.67	93.56	93.65	93.52
NB	82.55	82.54	82.54	82.64
DT	93.65	93.6	93.57	93.65
RF	95.89	95.88	95.84	95.68
XGBOOST	94.45	93.86	93.25	93.66

After training the dataset, we tried every potential strategy. Incorporate criteria for quality control and accuracy as well. In order to find the F1 score, precision, accuracy, and recall, we compared the predicted values with the test data. In Table 2 you can see the F1 score, recall, accuracy, and precision of the first model that predicted the assault plan. With a prediction accuracy of 95.55%, SVML was victorious in the data comparison. Just barely, the SVML algorithm beat out the competition, which included RF, LR, XGBoost, SVMK, DT, KNN, and NB. The success percentage was 82.54% in New Brunswick, which was the lowest. Similar results were obtained by other algorithms as NB. As illustrated in Figure 5A, the distribution graph of the observed values and the values predicted by the SVML technique is shown, while Figure 5B shows the error matrix. When comparing the model's accuracy, recall, and F1 scores, the SVML technique once again came out on top, however it was only by a little margin. You can get results over 93% with any of these models: LR, SVMK, DT, RF, or XGBoost. They all performed around the same. Scores were 9% below average for the worst-performing KNN and NB. On the whole, the results produced by each algorithm met expectations. These findings shown that machine learning can accurately forecast the trajectory of a cyberattack. Based on a person's profile, the proposed method would let users anticipate which crimes that person may encounter. Incorporate early warning systems as well. When comparing the model's accuracy, recall, and F1 scores, the SVML technique once again came out on top, however it was only by a little margin. You can get results over 93% with any of these models: LR, SVMK, DT, RF, or XGBoost. They all performed around the same. Compared to the other groups, KNN and NB had a 10% lower score, indicating their poor performance. On the whole, the results produced by each algorithm met expectations. These findings shown that machine learning can accurately forecast the trajectory of a cyberattack. You may enter a person's traits into the suggested model (Table 2), and it will forecast the sorts of crimes that person will commit.



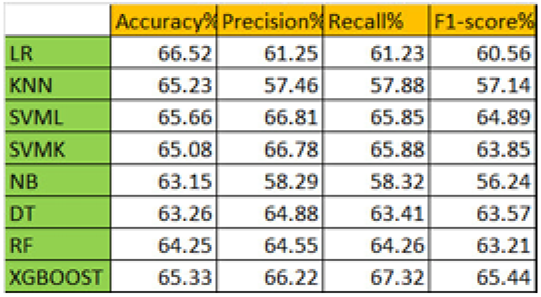
(A)



89	0	3	0	0	0
0	13	0	0	0	0
0	0	14	0	0	3
0	0	0	7	0	0
0	0	0	0	23	0
3	0	1	0	0	33

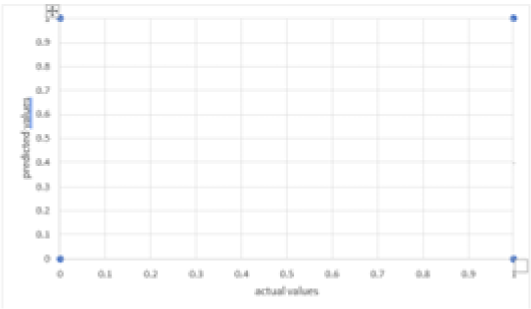
(B)

Figure 5: (A) 1st model comparison with actual values (B) Confusion matrix Table 3: Model 2 performance in machine learning model

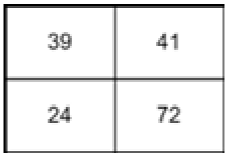


	Accuracy%	Precision%	Recall%	F1-score%
LR	66.52	61.25	61.23	60.56
KNN	65.23	57.46	57.88	57.14
SVML	65.66	66.81	65.85	64.89
SVMK	65.08	66.78	65.88	63.85
NB	63.15	58.29	58.32	56.24
DT	63.26	64.88	63.41	63.57
RF	64.25	64.55	64.26	63.21
XGBOOST	65.33	66.22	67.32	65.44

the following algorithms were used to get there: LR (66.52%), SVML (0.827%), KNN (1.44%), SVMK (1.39%), XGBoost (2.44%), RF (3.34%), and DT (3.34%). All of the algorithms were rather similar in their outcomes, with the exception of NB, which did the poorest. Figure 6B displays the distribution graph of the actual values and the projected values achieved using the SVML approach, while Figure 6A illustrates the error matrix. All of the algorithms were rather similar in their outcomes, with the exception of NB, which did the poorest. Here we can see the error matrix in Figure 6B and the distribution graph of the actual and projected values produced by the SVML algorithm in Figure 6A. The algorithms' output showed recall, accuracy, and F1 scores between 56% and 66%. Poor quality was the final product. We aimed to establish whether the same criminal was responsible for the crime by comparing the known and unknown features of the attacker. Still, it was recommended to build a new model with more characteristics based on the model's output.



(A)



39	41
24	72

(B)

Image 6: (A) Matrix of comparison inaccuracy (B) Confusion The suggested research study is constrained by the dataset's size, which contains genuine data. Temporal data may be used to estimate time series, but we must this data in order to proceed. In a similar vein, precise estimates can help discover the perpetrator if the technical details of the

Conclusion:

Direction in the Face of the COVID-19 Financial Pandemic, Frontiers in Psychology, Volume 12, Issue 12, 2022, Article No. 1664-1078, Combining machine learning algorithms with historical data on comparable efforts may help detect and halt cyberattacks, according to one research. The model predicts the traits of potential victims and the kinds of attacks they may face. The methods used by machine learning are compelling enough. Using linear SVMs is the way to go. With a 61% success rate, the algorithm can identify the cybercriminal responsible for an attack. Using a variety of AI methods, I propose increasing this figure. More people need to know that malware and social engineering are serious problems. Victims' income and education levels were negatively associated to the probability of a cyberattack. The primary goal of the research is to provide law enforcement with more efficient resources to combat cybercrime more aggressively. It is possible to develop new forms of training and warning systems for people with comparable characteristics by studying the characteristics of assault victims that surfaced during our investigation.

References:

[1]. Bilen, Abdulkadir & Özer, Ahmet. (2021). Cyber-attack method and perpetrator prediction using machine learning algorithms. PeerJ Computer Science.

[2]. 7. e475. 10.7717/peerj- cs.475. Al-majed, Rasha & Ibrahim, Amer &Abualkishik, Abedallah & Mourad, Nahia & Almansour, Faris. (2022). Using machine learning algorithm for detection of cyber-attacks in cyber physical systems. Periodicals of Engineering and Natural Sciences (PEN). 10. 261. 10.21533/pen.v10i3.3035.

[3]. Mazhar, T.; Irfan, H.M.; Khan, S.; Haq, I.; Ullah, I.; Iqbal, M.; Hamam, H. Analysis of Cyber Security Attacks and Its Solutions for the Smart grid Using Machine Learning and Blockchain Methods. Future Internet 2023, 15, 83. https://doi.org/10.3390/fi15020083

[4]. Sarker, I.H. Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. Ann. Data. Sci. https://doi.org/10.1007/s40745-022-00444-2 5. (2022). A. Alshehri, N. Khan, A. Alowayr and M. Yahya Alghamdi, "Cyberattack detection framework using machine learning and user behavior analytics," Computer Systems Science and Engineering, vol. 44, no.2, pp. 1679–1689, 2023.

[5]. Agrawal, K. K. ., P. . Sharma, G. . Kaur, S. . Keswani, R. . Rambabu, S. K. . Behra, K. . Tolani, and N. S. . Bhati. "Deep Learning-Enabled Image Segmentation for Precise Retinopathy Diagnosis". *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 12s, Jan. 2024, pp. 567-74, https://ijisae.org/index.php/IJISAE/article/view/4541.

[6]. Samota, H. ., Sharma, S. ., Khan, H. ., Malathy, M. ., Singh, G. ., Surjeet, S. and Rambabu, R. . (2024) "A Novel Approach to Predicting Personality Behaviour from Social Media Data Using Deep Learning", *International Journal of Intelligent Systems and Applications in Engineering*, 12(15s), pp. 539–547. Available at: https://ijisae.org/index.php/IJISAE/article/view/4788

[7].

[8]. Amjad Rehman, Tanzila Saba, Muhammad Zeeshan Khan, Robertas Damaševičius, Saeed Ali Bahaj, "Internet-of-Things Based Suspicious Activity Recognition Using Multimodalities of Computer Vision for Smart City Security", Security and Communication Networks, vol. 2022, Article ID 8383461, 12 pages, 2022. https://doi.org/10.1155/2022/8383461

[9]. Liu Qiang, Qu Xiaoli, Wang Dake, Abbas Jaffar, Mubeen Riaqa, Product Market Competition and Firm Performance: Business Survival Through Innovation and Entrepreneurial 10.3389/fpsyg.2021.790923. URL=https://www.frontiersin.org/articles/10.3389/fpsyg.2021. 790923

[10]. Ibor, A.E., Oladeji, F.A., Okunoye, O.B. et al. Conceptualisation of Cyberattack prediction with deep learning. Cybersecur 3, 14 (2020). https://doi.org/10.1186/s42400-020-00053-7 Yirui Wu, Dabao Wei, Jun Feng, "Network Attacks Detection Methods Based on Deep Learning Techniques: A Survey", Security and Communication Networks, vol. 2020, Article ID 8872923, 17 pages, 2020. https://doi.org/10.1155/2020/8872923

[11]. 10. Tehseen Mazhar, Hafiz Muhammad Irfan, Sunawar Khan, Inayatul Haq, Inam Ullah, Muhammad Iqbal, Habib

Hamam, Analysis of Cyber Security Attacks and Its Solutions for the Smart grid Using Machine Learning and Blockchain Methods, Future Internet, 10.3390/fi15020083, 15, 2, (83), (2023).

[12]. 11. McCarthy A, Ghadafi E, Andriotis P and Legg P. (2023). Defending against adversarial machine learning attacks using hierarchical learning. Journal of Information Security and Applications. 72:C.

[13]. 12. Ahsan, M.; Nygard, K.E.; Gomes, R.; Chowdhury, M.M.; Rifat, N.; Connolly, J.F. Machine Learning Techniques in Cybersecurity. Encyclopedia. Available online: https://encyclopedia.pub/entry/25675 (accessed on 30 April 2023).

[14]. Kenfack, P.D.B., Mbakop, F.K. and Eyong-Ebai, E. (2021) Implementation of Machine Learning Method for the Detection and Prevention of Attack in Supervised Network. Open Access Library Journal, 8, 1-25. doi: 10.4236/oalib.1108000.

[15]. Tewari, Shiv Hari, Data Science and Its Application in Cyber Security (Cyber Security Data Science) (September 5, 2020). Data Science in Cyber Security and cyber threat intelligence:

[16]. Sikos, Leslie F, Choo. Kim kwang. [Upcoming challenges in Cyber Security Data Science]. Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools: Tariq Mahmood , Uzma Afzal [Definition of Cyber, Available at SSRN: https://ssrn.com/abstract=3687251

[17]. Zhao L, Zhu D, Shafik W, et al. Artificial intelligence analysis in cyber domain: A review. International Journal of Distributed Sensor Networks. 2022;18(4). doi:10.1177/15501329221084882

[18]. Narayan, Valliammal, and Barani Shaju. "Malware and Anomaly Detection Using Machine Learning and Deep Learning Methods." Research Anthology on Machine Learning Techniques, Methods, and Applications, edited by Information Resources Management Association, IGI Global, 2022, pp. 149-176. https://doi.org/10.4018/978-1-6684-6291-1.ch010

[19]. Ahmad Naim Irfan, SuriayatiChuprat, Mohd Naz'riMahrin, Aswami Ariffin. (2022) Taxonomy of Cyber Threat Intelligence Framework. 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), pages 1295-1300.

[20]. Aksu, Dogukan, and M. Ali Aydin. "Detecting port scan attempts with comparative analysis of deep learning and support vector machine algorithms." 2018 International congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT). IEEE, 2018.

[21]. Khuphiran, Panida, et al. "Performance comparison of machine learning models for DDoS attacks detection." 2018 22nd International Computer Science and Engineering Conference