Scientific and Practical Cyber Security Journal (SPCSJ) 8(1): 64 – 68 ISSN 2587-4667 Scientific Cyber Security Association (SCSA)

კიბერ ინციდენტების პროგნოზირება მანქანური სწავლების ალგორითმების გამოყენებით

PREDICTING CYBER INCIDENTS USING MACHINE LEARNING ALGORITHMS

Tinatin Mshvidobadze[1]
[1]Gori State University, Gori, Georgia

აბსტრაქტი. ნაშრომში წარმოდგენილია კიბერ ინციდენტებთან დაკავშირებული მეთოდები, სხვადასხვა მკვლევარების მიერ. მანქანური სწავლების ალგორითმები (DM-ML) მნიშვნელოვან როლს თამაშობს კიბერუსაფრთხოების[1] სფეროში კიბერ ინციდენტების (SCI) პროგნოზირებასა და გამოვლენაში. ნაშრომში მოცემულია კარგად ცნობილია ML კლასიფიკატორები, მონაცემთა კლასიფიკაციისათვის. მონაცემები აღებულია სტრატეგიული და საერთაშორისო კვლევების ცენტრის (CSIS) ანგარიშის მიხედვით. განხილულია ცენტრალიზებული კლასიფიკატორის მიდგომა მსოფლიოს ექვსი კონტინენტის მონაცემების მიხედვით. ნაშრომში კლასიფიკატორების შედარების საფუძველზე მაღალი სიზუსტით პროგნოზირებულია, თუ რომელი ტიპის კიბერ ინციდენტი შეიძლება მოხდეს და მსოფლიოს რომელ ნაწილში.

საკვანძო სიტყვები: კიბერ ინციდენტი, კიბერუსაფრთხოება, მონაცემთა მოპოვება, მანქანური სწავლება.

ABSTRACT. The paper presents methods related to cyber incidents by various researchers. Machine learning algorithms (DM-ML) play an important role in the prediction and detection of cyber incidents (SCI) in the field of cyber security. The paper presents well-known ML classifiers for data classification. The data set is taken from a report by the Center for Strategic and International Studies (CSIS). A centralized classifier approach based on data from six continents of the world is discussed. Based on the comparison of classifiers in the paper, it is predicted with high accuracy which type of SCI may occur and in which part of the world.

KEYWORDS: cyber incidents, cyber security, data mining, machine learning.

1.შესავალი

IoT და 5G ტექნოლოგიების სწრაფი ზრდა კიბერსივრცეს არაუსაფრთხოს ხდის, რაც საბოლოოდ იწვევს მნიშვნელოვანი კიბერ ინციდენტების განვითარებას [1]. მოსალოდნელია, რომ IoT მოწყობილობების რაოდენობა 2025 წლისთვის დაახლოებით 75 მილიარდს მიაღწევს

[1] კიბერუსაფრთხოება არის ტექნიკა, რომელიც იცავს სისტემას ინტერნეტში კიბერ ინციდენტებისაგან.

[2]. "კიბერუსაფრთხოების ალმანახის" მიხედვით, რომელიც გამოქვეყნდა "Cybersecurity Ventures"-ის მიერ, გლობალური კიბერდანაშაულის ღირებულება 2025 წელს 10,5 ტრილიონ აშშ დოლარს მიაღწევს.

კიბერ ინციდენტი ნიშნავს აქტივობას ან მოვლენას, რომელიც ხდება ინტერნეტის საშუალებით და საფრთხეს უქმნის საკომუნიკაციო სისტემის კონფიდენციალურობას, მთლიანობასა და ხელმისაწვდომობას ნებისმიერი საშუალებით. ტერმინი მნიშვნელოვანი კიბერ ინციდენტი (SCI) ნიშნავს ინციდენტს, რომელიც იწვევს ეროვნული უსაფრთხოებისა და ეკონომიკის აშკარა ზიანს [3].

SCI-ის ზრდასთან ერთად, კიბერუსაფრთხოების ზომები ასევე გაუმჯობესდა ამ ინციდენტების მოსაგვარებლად. მონაცემთა მოპოვება და მანქანური სწავლება (DM-ML) მნიშვნელოვან როლს თამაშობს კიბერ ინციდენტების პროგნოზირებაში, პრევენციასა და გამოვლენაში სხვადასხვა მიდგომების გამოყენებით [4].

ნაშრომში განვიხილავთ სხვადასხვა მკვლევარების მიერ მიღებულ ეფექტურ შედეგებს ამ ინციდენტების აღმოსაფხვრელად.

მოცემულია უსმან აშრაფის და სხვა მკვლევარების მიერ პროექტის ფარგლებში ჩატარებული კვლევის შედეგები [5]. ასევე ნაჩვენებია ცენტრალიზებული კლასიფიკატორის სარგებელი მომავალში *SCI* -ის აღმოსაფხვრელად [2].

ML ალგორითმები, როგორიცაა ნაივ ბაიესი (NB) [6], დამხმარე ვექტორული აპარატი (SVM) [7], ლოგისტიკური რეგრესია (LR)[8] და გადაწყვეტილებათა ტყე (RF)[9] გამოიყენება მონაცემთა კლასიფიკაციისათვის [10], კიბერ ინციდენტების პრევენციასა და პროგნოზირებისათვის.

2. ლიტერატურის მიმოხილვა

კიბერუსაფრთხოება არის განვითარებადი და უზარმაზარი გამოწვევა მსოფლიოში სხვადასხვა კიბერ ინციდენტებთან დაკავშირებით. მნიშვნელოვანი ნაწილია არსებული ინციდენტების იდენტიფიცირება სხვადასხვა DM-ML ალგორითმის გამოყენებით. DM-ML-ზე დაფუძნებული მიდგომები არის ძალიან ცნობილი ტექნიკა, რომლებიც გამოიყენება კიბერუსაფრთხოების დაუცველობის გამოსავლენად და სწორედ ამიტომ გამოიყენება BoW მოდელში, ხოლო კლასიფიკატორისათვის გამოიყენება NB, SVM, LR და RF ალგორითმები.

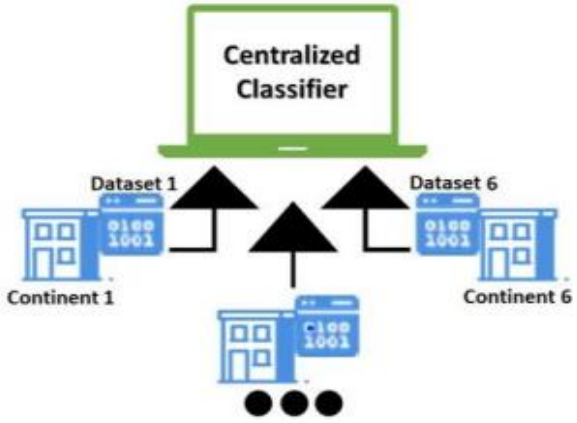
ბისვასმა და სხვა ავტორებმა [11] გამოიყენეს ტექსტის მოპოვების მიდგომა ციფრული ჯანდაცვის სფეროში კიბერ ინციდენტების გამოსავლენად. ავტორებმა გამოიყენეს ბუნებრივი ენის დამუშავება (NLP) ახალი ამბების მონაცემების მოსაპოვებლად და ინფორმაციის მისაღებად.

სურიმ და სხვა ავტორებმა [12] წარმოადგინეს ანომალიის დეტექტორი ავარიის შეტყობინების გამოყენებით. ისინი მუშაობდნენ ტექსტურ მონაცემებზე და გამოიყენეს *Local Outlier Factor* (LoF) ანომალიური მდგომარეობის გამოსავლენად. ავტორებმა გამოიკვლიეს სხვადასხვა DM-ML მიდგომები მავნე პროგრამების აღმოსაჩენად, ასევე ღრმა სწავლების მეთოდოლოგია, რომელიც გამოიყენება კიბერშეტევების პროგნოზირებისათვის, ქსელის ტრაფიკიდან მიღებული მონაცემების საფუძველზე.

[2] Research work through the project number: IFP22UQU4310108DSR188.

ფანგმა და სხვებმა [13], შეიმუშავეს კიბერშეტევის მეთოდები დამხმარე ვექტორის აპარატის (SVM) გამოყენებით, ML ალგორითმში. ავტორებმა დაასკვნეს სხვადასხვა DM-ML მიდგომები, როგორიცაა ბაიესის ქსელი, გადაწყვეტილების ხე, კლასტერირება და ხელოვნური ნეირონული ქსელები (ANN) კიბერუსაფრთხოებაში კიბერ ინციდენტების გამოსავლენად.

აშრაფმა და სხვა მკვლევარებმა აჩვენეს ცენტრალიზებული კლასიფიკატორის გამოყენების ეფექტურობა (ნახ.1). ნაჩვენებია, თუ რომელი ტიპის *SCI* მოხდა და მსოფლიოს რომელ კონტინენტზე, მონაცემთა ნაკრები გამოიყენეს ცენტრალიზებული კლასიფიკატორის მოსამზადებლად თითოეული კონტინენტისათვის.



სურათი 1. ცენტრალიზებული კლასიფიკატორის მონაცემთა ბაზა ექვსი კონტინენტის მიხედვით

მონაცემთა ნაკრები არის *SCI*-ის ტიპი, რომელიც მოხდა მსოფლიოს 6 კონტინენტზე (2003 წლის სექტემბრიდან 2023 წლის ოქტომბრამდე), სტრატეგიული და საერთაშორისო კვლევების ცენტრის (CSIS) ანგარიშის მიხედვით[14]. SCI რაოდენობა აზიისთვის უფრო მაღალია, რადგან ეს არის ყველაზე დიდი კონტინენტი მსოფლიოში.(ცხ.1.).

ცხრილი I
მონაცემების განაწილება

SCI ტიპი	აფრიკა	აზია	ევროპა	ჩრდილოეთ ამერიკა	ოკეანია	სამხრეთ ამერიკა	SCI რაოდენობა
APT	4	62	25	20	5	0	116
DDoS	0	15	13	6	3	0	37
DoS	0	1	0	0	0	0	1
Espionage	1	8	7	7	1	0	24
Malware	4	46	53	19	3	0	125
Man-in-Middle	1	1	1	1	1	0	5
Phishing	3	55	86	47	6	8	205
SQL Injection	2	3	6	2	0	0	13
Total	15	203	194	103	20	8	543

ამ კვლევამ გამოავლინა, გამოიკვლია და გადაჭრა, თუ როგორ უნდა გამოვთვალოთ კონტინენტის სახელი *SCI*-ის ტიპის მიხედვით.

კლასიფიკაციისათვის გამოიყენება მანქანური სწავლების ოთხი განსხვავებული კლასიფიკატორი [15]:

ნაივ ბაიესი (NB) - იგი ეფუძნება ბაიესის თეორემას, რომელიც გამომდინარეობს პირობითი ალბათობიდან. ის ჩვეულებრივ გამოიყენება ზედამხედველობით სწავლაში ტექსტის მონაცემთა კლასიფიკაციისთვის. *NB* ეფექტურია არაწრფივი ამოცანებისთვის.

დამხმარე ვექტორის აპარატი (SVM) - ეს არის ზედამხედველობითი სწავლebის კლასიფიკატორი. *SVM* არის ვექტორული მიდგომა და ძალიან ეფექტურია, თუ პრობლემა წრფივია და მონაცემთა ნაკრები შეზღუდულია.

ლოგისტიკური რეგრესია (LR) - პროგნოზირებს ორობით პრობლემას და მის შედეგებს ეფექტურად. ის გვაწვდის ინფორმაციას მახასიათებლების სტატისტიკური მნიშვნელობის შესახებ და იყენებს ალბათურ მიდგომას.

გადაწყვეტილებათა ტყე (RF) - *Random Forest* შედგება მრავალი გადაწყვეტილების ხისგან, მოდელის ეფექტურობის გაზრდით. ის ასევე მუშაობს არაწრფივ ამოცანებზე. ტექნიკური თვალსაზრისით, ეს არის მეთოდი გადაწყვეტილების ხეების გენერირებისათვის მონაცემთა ნაკრების ქვეჯგუფიდან.

პროექტის შესრულებისას გამოიყენეს უნიგრამის და ბიგრამის მოდელების კონცეფცია, მონაცემებიდან სიტყვების გასაფილტრად მინიმალური სიხშირით.

ექსპერიმენტული კვლევისას კლასიფიკატორების გამოყენების შედეგია კონტინენტის სახელის პროგნოზირება *SCI*-ის ტიპის მიხედვით. კლასიფიკატორების მუშაობის შესაფასებლად, გამოიყენება სიზუსტე და F1-ზომა, როგორც შესრულების ინდიკატორები. სიზუსტის ზომები NB, LR და RF არის (0.952396, 0.920829), (0.984139, 0.962375), (0.978099, 0.962375) შესაბამისად. SVM, NB, LR და RF კლასიფიკატორები შეფასდა სათითაოდ და დადგინდა რომ აზია, ყველაზე მეტად დაზარალებული რეგიონია *SCI* კუთხით.

3. დასკვნა

ეს ნაშრომი ფოკუსირებულია 2003 წლის სექტემბრიდან 2023 წლის ოქტომბრის ჩათვლით მნიშვნელოვან კიბერ ინციდენტებზე (SCI) დაფუძნებულ კვლევაზე, სტრატეგიული და საერთაშორისო კვლევების ცენტრის (CSIS) ანგარიშის მიხედვით. ოთხი განსხვავებული კლასიფიკატორით, ასევე პროგნოზირებულია რომელი კონტინენტი უფრო მეტად განიცდის *SCI*-ს ამ პერიოდის განმავლობაში.

მომავალში, *SCI*-სათვის სხვადასხვა მონაცემთა ნაკრები შეიძლება განიხილებოდეს და სხვადასხვა მანქანური სწავლების კლასიფიკატორების გამოყენებით, მათი ეფექტურობის შესამოწმებლად, როგორიცაა ფედერალური მანქანური სწავლება (FML). გარდა ამისა, აღნიშნულ მოდელში უსაფრთხოების გასაძლიერებლად ასევე შეიძლება Blockchain-ის განხორციელება.

გამოყენებული ლიტერატურა:

1. Li Y., and Liu Q., 2021, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments," Energy Reports, vol. 7, pp. 8176–8186;
2. Hejase H., Kazan H., Hejase A, and Moukadem I, 2021, "Hejase et al. Cyber Security paper," Computer and Information Science, vol. Vol. 14, pp. 10–25, doi: 10.5539/cis. v14n2p10;
3. Hodgson Q,, Clark-Ginsberg A., Haldeman Z., Lauland, A and Mitch I.,2022, Managing Response to Significant Cyber Incidents: Comparing Event Life Cycles and Incident Response Across Cyber and Non Cyber Events. Santa Monica, CA: RAND Corporation, doi: 10.7249/RRA1265-4;
4. Handa A., Sharma A., and Shukla S., 2019, "Machine learning in cybersecurity: A review," WIREs Data Mining and Knowledge Discovery, vol. 9, no. 4, p. e1306, doi: 10.1002/widm.1306;
5. Mumtaz G., Akram S., Waseem M., Iqbal M., Ashraf U., Almarhabi K., Mohammed A., and Adel A., 2017, "Classification and Prediction of Significant Cyber Incidents (SCI) using Data Mining and Machine Learning (DM-ML)".
6. Alqahtani H,. Sarker I., Kalim, A., Minhaz Hossain M., Ikhlaq S., and Hossain 2020, "Cyber Intrusion Detection Using Machine Learning Classification Techniques," in Computing Science, Communication and Security, Singapore, pp. 121–131.;
7. Bhusal N., Gautam M., and Benidris M., 2021, "Detection of Cyber Attacks on Voltage Regulation in Distribution Systems Using Machine Learning," IEEE Access, vol. 9, pp. 40402–40416, doi: 10.1109/ACCESS.2021.3064689.
8. Bapat R., et al., 2018, "Identifying malicious botnet traffic using logistic regression," in Systems and Information Engineering Design Symposium (SIEDS), pp. 266–271. doi: 10.1109/SIEDS.2018.8374749;
9. Ustebay S., Turgut Z., and Aydin M., "Intrusion Detection System with Recursive Feature Elimination by Using Random Forest and Deep Learning Classifier," in 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Dec. 2018, pp. 71–76. doi: 10.1109/IBIGDELFT.2018.8625318.;
10. Chayal N., and Patel N., 2021, "Review of Machine Learning and Data Mining Methods to Predict Different Cyberattacks," in Data Science and Intelligent Applications, Singapore, pp. 43–51;
11. Biswas B., Mukhopadhyay A., Bhattacharjee S., Kumar A., and Delen D., 2022, "A text-mining based cyber-risk assessment and mitigation framework for critical analysis of online hacker forums," Decision Support Systems, vol. 152, p. 113651, doi: 10.1016/j.dss.2021.113651;
12. Souri A., and Hosseini R., 2018, "A state-of-the-art survey of malware detection approaches using data mining techniques," Hu-man-centric Computing and Information Sciences, vol. 8, no. 1, p. 3, doi: 10.1186/s13673-018-0125-x;
13. Fang X., Xu M., and Zhao P., 2019, "A deep learning framework for predicting cyber-attacks rates," EURASIP Journal on Information Security, doi: 10.1186/s13635-019-0090-6;
14. "Significant Cyber Incidents (SCIs)." [Online]. Available: https://www.csis.org/programs/strategictechnologies-program/significant-cyber-incidents;
15. Xu S., 2018, "Bayesian Naïve Bayes classifiers to text classification," J. Inf. Sci., vol. 44, no. 1, pp. 48–59, doi: 10.1177/0165551516677946.