International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1006-1010

Artificial Intelligence for Cyber Security: A Systematic Mapping of Literature

A.Uday Kiran [1], S.Neha [2], A.Manasa [3], N.Anusha [4]
[1]*Assistant Professor, CMR Technical Campus,*
[2,3,4]*CMR Technical Campus*

ABSTRACT
Due to the ever-increasing complexities in cybercrimes, there is the need for cybersecurity methods to be more robust and intelligent. This will make defense mechanisms to be capable of making real-time decisions that can effectively respond to sophisticated attacks. To support this, both researchers and practitioners need to be familiar with current methods of ensuring cybersecurity (CyberSec). In particular, the use of artificial intelligence for combating cybercrimes. However, there is lack of summaries on artificial intelligent methods for combating cybercrimes. To address this knowledge gap, this study sampled 131 articles from two main scholarly databases (ACM digital library and IEEE Xplore). Using a systematic mapping, the articles were analyzed using quantitative and qualitative methods. It was observed that artificial intelligent methods have made remarkable contributions to combating cybercrimes with significant improvement in intrusion detection systems. It was also observed that there is a reduction in computational complexity, model training times and false alarms. However, there is a significant skewness within the domain. Most studies have focused on intrusion detection and prevention systems, and the most dominant technique used was support vector machines. The findings also revealed that majority of the studies were published in two journal outlets. It is therefore suggested that to enhance research in artificial intelligence for CyberSec, researchers need to adopt newer techniques and also publish in other related outlets.

Keywords: Artificial intelligence and cybersecurity, information security, machine learning, systematic reviews.

Date of Submission: 02-06-2022 Date of acceptance: 16-06-2022

I. INTRODUCTION

The rapid evolution of information and communication technologies, including the Internet has bred positive implications to organizations and social lives. The Internet provides a platform that facilitates communication and networking. It supports knowledge sharing [1] and social interaction [2] which are important ingredients for human development. Amidst all the benefits, it has a dark side. Its increase in reliance on third party and/or cloud-based data storage and applications, make it extremely difficult for organizations to provide ‘‘total’’ security to their information systems.

For instance, current cloud infrastructure is characterized by a three-layer architecture. Each layer is arguably at risk from a range of vulnerabilities introduced either by programmers or service providers. The disparate handling of data makes crimes such as cybertheft and cyber-fraud more complex to track within cyberspaces [3]. More worryingly, ubiquitous distributed computing eliminates the importance of geographical boundaries, and this also makes it possible for criminal activities to originate from any part of the World [4]. Hence, organizations are increasingly challenged by a wide range of cyber-attacks [5], [6].

These attacks are characterized by a high level of sophistication that calls for tthe need of adopting Artificial Intelligence (AI) or intelligent agents to combat them.

Accordingly, cyber defense mechanisms must be i) increasingly intelligent, ii) more flexible, and iii) robust enough to detect a variety of threats and mitigate against them. To achieve these requirements, organizations are adopting AI techniques to effectively monitor and combat cyber-attacks and cybercrimes. This, in addition, calls for the need for researchers and practitioners to be familiar with current state of the art on the use of AI methods for cyber safety. Although some existing studies have summarized and discussed issues impacting cybersecurity (CyberSec), to the best of our knowledge none has, in a systematic manner, focused on AI applications in CyberSec. Thus, this paper aims to systematically review existing studies on the use of AI techniques for combating cyber-attacks.

The next section provides a brief background on how AIs are used in combating CyberSec issues. This is followed by a discussion on existing related works and the current knowledge gap. The method used for conducting the study, the findings, discussions and conclusions are also presented.

II. LITERATURE SURVEY

A.Intrusion detection systems

Intrusion can be defined as any kind of unauthorised activities that cause damage to an information system. This means any attack that could pose a possible threat to the information confidentiality, integrity or availability will be considered an intrusion. For example, activities that would make the computer services unresponsive to legitimate users are considered an intrusion. An IDS is a software or hardware system that identifies malicious actions on computer systems in order to allow for system security to be maintained. The goal of an IDS is to identify different kinds of malicious network traffic and computer usage, which cannot be identified by a traditional firewall.

B.Signature-based intrusion detection systems

Signature intrusion detection systems (SIDS) are based on pattern matching techniques to find a known attack; these are also known as Knowledge-based Detection or Misuse Detection (Khraisat et al., 2018). In SIDS, matching methods are used to find a previous intrusion. In other words, when an intrusion signature matches with the signature of a previous intrusion that already exists in the signature database, an alarm signal is triggered. For SIDS, host's logs are inspected to find sequences of commands or actions which have previously been identified as malware. SIDS have also been labelled in the literature as Knowledge-Based Detection or Misuse Detection.

C.Anomaly-based intrusion detection system (AIDS)

AIDS has drawn interest from a lot of scholars due to its capacity to overcome the limitation of SIDS. In AIDS, a normal model of the behavior of a computer system is created using machine learning, statistical-based or knowledge-based methods. Any significant deviation between the observed behavior and the model is regarded as an anomaly, which can be interpreted as an intrusion. The assumption for this group of techniques is that malicious behavior differs from typical user behavior.

D.Python

Python is a programming language, which means it'a a language both people and computers can understand. Python was developed by a Dutch software engineer named Guido van Rossum, who created the language to solve some problems he saw in computer languages of the time.Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released program. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

III. PROPOSED SYSTEM

Faced with imbalanced network traffic data, we propose a novel Difficult Set Sampling Technique(DSSTE) algorithm to tackle the class imbalance problem in network traffic. This method effectively reduces the imbalance and makes the classification model learning difficult samples more effective. We use classic machine learning and deep learning algorithms to verify on two benchmark datasets. The specific contributions are as follows. (1) We use the classic NSL-KDD and the up-to-date CSECIC-IDS2018 as benchmark datasets and conduct detailed analysis and data cleaning. (2)This work proposes a novel DSSTE algorithm,reducing the majority samples and augmenting the minority samples in the difficult set, tackling the class imbalance problem in intrusion detection so that the classifier learns the differences better in training. (3) The classification model uses Random Forest(RF), Support Vector Machine(SVM), .

IV. METHODOLOGY

A. System Architecture

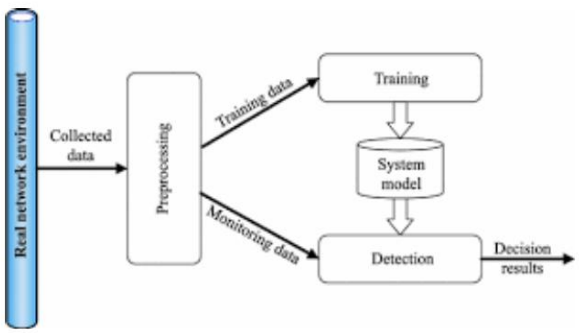


Fig 1: BLOCK DIAGRAM

B.Modules

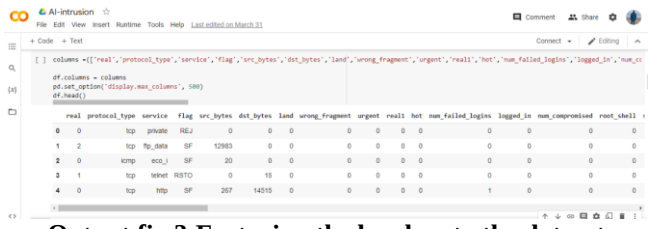
Data Source KDD CUP dataset is used for the experiment. It has 42 attributes classified under numeric, nominal, and binary. Two types of classes are there Normal and Anomaly. Anomaly class has some attacks categorized as dos, probe, u2r, r2l. This dataset has enough volume of records for training and testing purposes.

- **Feature Extraction:** Feature selection is made in a dataset to lessen the computational effort and find the optimal subset of the dataset that produces higher classification accuracy. A few parts if a dataset include irrelevant and redundant features, and such features distract the classifiers. PSO is used in this experiment for the extraction of features. 9 relevant features are selected among 42 features of the KDD-CUP dataset. .

OUTPUT

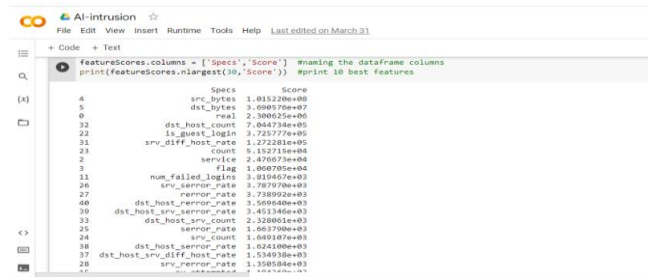
Output fig 1: Loading the libraries required for the model

Output fig 2:Loading the data and seeing the few lines of the dataset



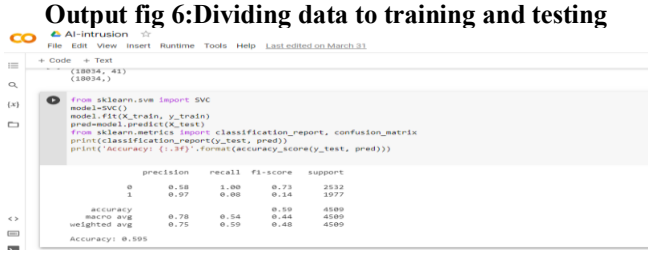
Output fig 3:Featuring the headers to the dataset

Output fig 4:Label encoder for converting data to numeric format

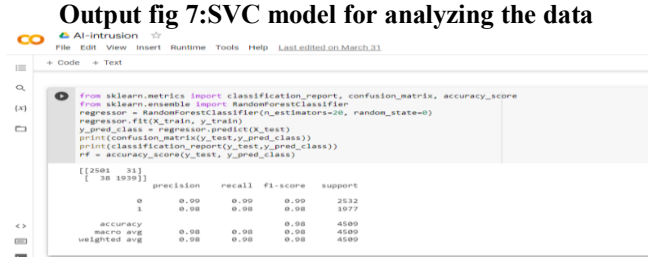


Output fig 5:Chi2 model for identifying the parameters

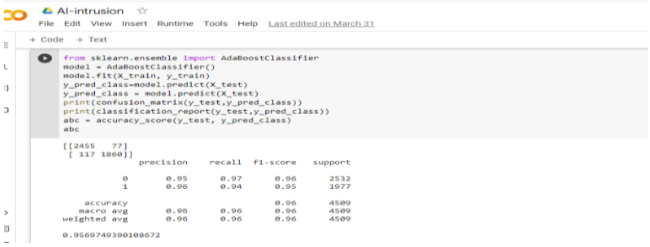
Output fig 6:Dividing data to training and testing



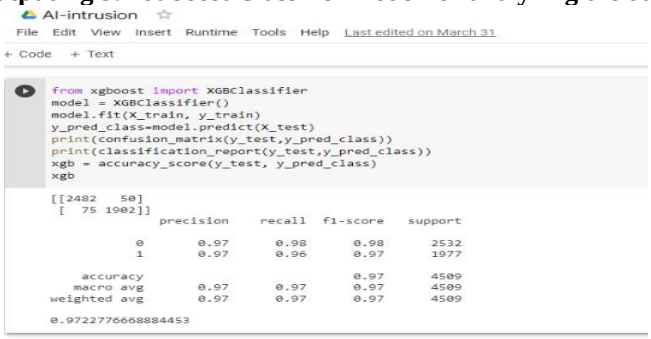
Output fig 7:SVC model for analyzing the data



Output fig 8:Random Forest model for analyzing the data



Output fig 9:Adaboost Classifier model for analyzing the data



Output fig 10:XGB Classifier for model

V. CONCLUSION AND FUTURE ENHANCEMENT

In conclusion, the study suggests that the application of AI in the CyberSec domain has been promising with IDPS showing improvement. AI has facilitated a reduction in computational complexity and reduced model training times. It was also observed that there is a considerable skewness within the domain. Moreover, researchers have focused on fewer algorithms and as such newer algorithms are not popular. This stands as both a challenge and also an opportunity for researchers. It is believed that AI applications will continue to offer opportunities for cybersecurity. However, research must never stand still, and researchers need to start adopting and adapting new approaches and publish more widely.

REFERENCES

[1]. D. E. Denning, ''An intrusion-detection model,'' IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

[2]. N. B. Amor, S. Benferhat, and Z. Elouedi, ''Naive Bayes vs decision trees in intrusion detection systems,''inProc.ACMSymp.Appl.Comput.(SAC), 2004, pp.420–424.

[3]. M. Panda and M. R. Patra, ''Network intrusion detection using Naive Bayes,''Int.J.Comput.Sci.Netw.Secur.,vol.7,no.12,pp. 258–263,2007.

[4]. M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, ''Support vector machine and random forest modeling for intrusion detection system (IDS),'' J. Intell. Learn.Syst. Appl., vol. 6, no. 1, pp. 45–52, 2014.

[5]. N. Japkowicz, ''The class imbalance problem: Significance and strategies,'' in Proc. Int. Conf. Artif. Intell., vol. 56, 2000, pp. 111–117.

[6]. Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[7]. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, ''Deep learning for visual understanding: A review,'' Neurocomputing, vol. 187, pp. 27–48, Apr. 2016.

[8]. T. Young, D. Hazarika, S. Poria, and E. Cambria, ''Recent trends in deep learning based natural language processing [review article],'' IEEE Comput. Intell. Mag.,vol. 13, no. 3, pp. 55–75, Aug. 2018

[9]. N.Shone,T.N.Ngoc,V.D.Phai,andQ.Shi, A deep learning network intrusion detection,'' IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp.41–50, Feb. 2018.

[10]. D. A. Cieslak, N. V. Chawla, and A. Striegel,''Combating imbalance in network intrusion datasets,'' in Proc. IEEE Int. Conf. Granular Comput., May 2006, pp.732–737.