International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Machine Learning and Deep Learning Approaches for Cyber Security

Mrs. Deepika Avinash Bhosale[1], Mr. Mahesh. J. Kanase[2]

[1,2]Professor, Department of Computer Science And Engineering, D.K.T.E. Society's YaswantraoChavan Polytechnic, Ichalkarnji, India

Abstract

Cyber security mainly prevents the hardware, software, and data present in the system that has an active internet connection from external attacks. Organizations mainly deploy cyber security for their databases and systems to prevent it from unauthorized access. This survey paper describes a Cyber security hiring and retention challenges are bigger than ever this year. View the State of Cyber security 2022 infographic to see the cyber workforce challenges and opportunities faced by enterprises around the world--and to see how your organization compares. The objective of this research work is to present the evaluation of some of the widely used machine learning techniques used to detect some of the most threatening cyber threats to the cyberspace. Three primary machine learning techniques are mainly investigated, including deep belief network, decision tree and support vector machine. We have presented a brief exploration to gauge the performance of these machine learning techniques in the spam detection, intrusion detection and malware detection starting from IP traffic classification, filtering malicious traffic for intrusion detection based on frequently used and benchmark datasets. Various attacks have been classified using the ML algorithms and finally the performance of each algorithm have been assessed. The complexity of ML/DM algorithms is addressed, discussion of challenges for using ML/DM for cyber security is presented, and some recommendations on when to use a given method are provided.

Keywords: Cyber Security, artificial intelligence, data mining, machine learning, intrusion detection, deep learning.

1. Introduction

Machine learning (ML) lets computers learn without being explicitly programmed. Put another way, machine learning teaches computers to do what people do: learn by experience. Machine learning is a domain within the broader field of artificial intelligence. In security, machine learning continuously learns by analyzing data to find patterns so we can better detect malware in encrypted traffic, find insider threats, predict where "bad neighborrhoods" are online to keep people safe when browsing, or protect data in the cloud by uncovering suspicious user behavior. The cyber threat landscape forces organizations to constantly track and correlate millions of external and internal data points across their infrastructure and users. It simply is not feasible to manage this volume of information with only a team of people. This is where machine learning shines, because it can recognize patterns and predict threats in massive data sets, all at machine speed. By automating the analysis, cyber teams can rapidly detect threats and isolate situations that need deeper human analysis. Machine learning detects threats by

constantly monitoring the behavior of the network for anomalies. Machine learning engines process massive amounts of data in near real time to discover critical incidents. These techniques allow for the detection of insider threats, unknown malware, and policy violations. Machine learning can predict "bad neighborhoods" online to help prevent people from connecting to malicious websites. Machine learning analyzes Internet activity to automatically identify attack infrastructures staged for current and emergent threats.

Algorithms can detect never-before-seen malware that is trying to run on endpoints. It identifies new malicious files and activity based on the attributes and behaviors of known malware. Machine learning can protect productivity by analyzing suspicious cloud app login activity, detecting location-based anomalies, and conducting IP reputation analysis to identify threats and risks in cloud apps and platforms. Machine learning can detect malware in encrypted traffic by analyzing encrypted traffic data elements in common network telemetry. Rather than decrypting, machine learning algorithms pinpoint malicious patterns to find threats hidden with encryption. Deep Learning (DL) is viewed as a specific set of machine learning approaches. Systems that use machine learning assist in making judgments based on data collected and automatically modify their model when discovering new patterns. It employs neural networks with numerous layers to learn from data. Deep learning is used when a model is trained for voice recognition, image recognition, and natural language processing applications for instance. The data processing chain has a complex structure since it goes through multiple layers where hidden inputs and outputs carry out prediction tasks and send the results to the next layer.

2. A METHODOLOGY

Method and analysis which is performed in your research work should be written in this section. A simple strategy to follow is to use keywords from your title in first few sentences. Cyber security in depth covers five important elements: detection, protection, management, response and recovery. Data mining is the process of analyzing information, discovering new patterns and data, and predicting future trends. It's often used in scientific research, business development, customer relations, and other spheres.

While the term data mining is usually treated as a synonym for knowledge discovery in databases (KDD), it's actually just one of the steps in the KDD process. One can view ML as the older sibling of DM. The term datamining was introduced in late 1980s (the first KDD conference took place in 1989), whereas the term machine learning has been in use since the 1960s. Presently, the younger sibling(i.e., use of the term DM) is more popular than the older one, which might be the reason why some researchers actually label their work as DM rather than ML. The main goal of KDD is to obtain useful and often previously unknown information from large sets of data. KDD is widely applied in any field that can benefit from the analysis of vast amounts of data: scientific studies, business analysis, marketing research, etc. It's also used by cybercriminals to find new ways of attacks, and by cyber security professionals to detect and stop these new attacks. Combining data mining and cyber security allows for determining features of cyber attacks and improving attack detection processes. To obtain valuable knowledge, data mining uses methods from statistics, machine learning (ML), artificial intelligence (AI), and database systems. ML is the one of the promising answers that can be effective against zero day threats. New research is being done by use of statistical traffic characteristics and ML techniques. This paper is a focused literature survey of machine learning and its application to cyber analytics for intrusion detection, traffic classification and applications such as email filtering. Based on the relevance

and the number of citation each methods were identified and summarized. Because datasets are an important part of the ML approaches some well know datasets are also mentioned. Some recommendations are also provided on when to use a given algorithm. An evaluation of four ML algorithms has been performed on MODBUS data collected from a gas pipeline. Various attacks have been classified using the ML algorithms and finally the performance of each algorithm have been assessed.

An ML method primarily includes the following four steps:

- Feature Engineering. Choice as a basis for prediction (attributes, features).
- Choose the appropriate machine learning algorithm.
- (Such as classification algorithm or regression algorithm, high complexity or fast)
- Train and evaluate model performance. (For different algorithms, evaluate and select the best performing model.)
- Use the trained model to classify or predict the unknown data.

Intrusion Detection

The first standard dataset provides a large amount of background traffic data and attack data. It can be downloaded directly from the website. Currently, the dataset primarily includes the following three data subsets:

- 1998 DARPA Intrusion Detection Assessment Dataset: Includes 7 weeks of training data and 2 weeks of test data.
- 1999 DARPA Intrusion Detection Assessment Dataset: Includes 3 weeks of training data and 2 weeks of test data.
- 2000 DARPA Intrusion Detection Scenario-Specific

Dataset: Includes LLDOS 1.0 Attack Scenario Data, LLDOS 2.0.2 Attack scenario data, Windows NT attack data.

INTRUSION DETECTION SYSTEMS IN RECENT WORKS USING MACHINE LEARNING AND DEEP LEARNING

Methodologies and algorithms have undergone significant change and evolution to produce the most acceptable intrusion detection system in many applications that attempt to identify constantly changing threats and attacks. Initially, classification was based on machine learning, but as performance needed to be further improved, deep learning was utilized to produce higher accuracy and a lower false alarm rate.

Therefore, the basic objective of this work is to introduce a bibliometric analysis of the deep learning approach used for the detection of potential threats to cyber security. Effectively, we have chosen the research papers from the year 2011 to 2020, which are based on cyber security issues with deep learning concepts

3. MODELING AND ANALYSIS

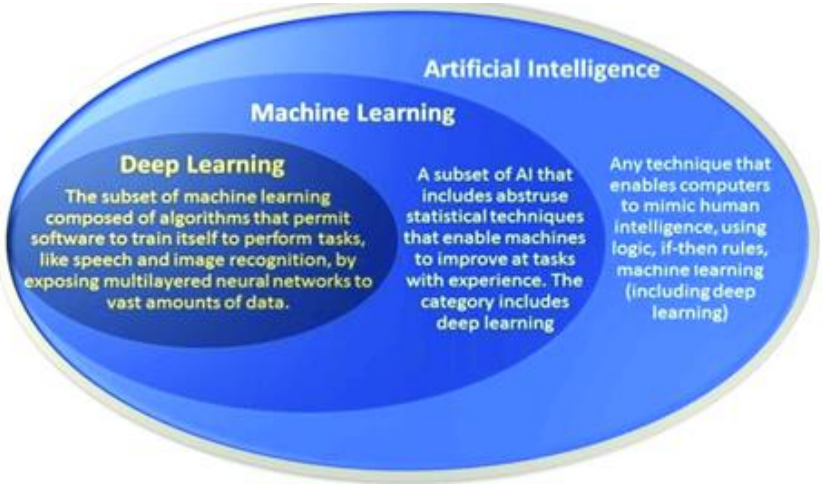

Figure 1: Relationship of AI, ML, and DL

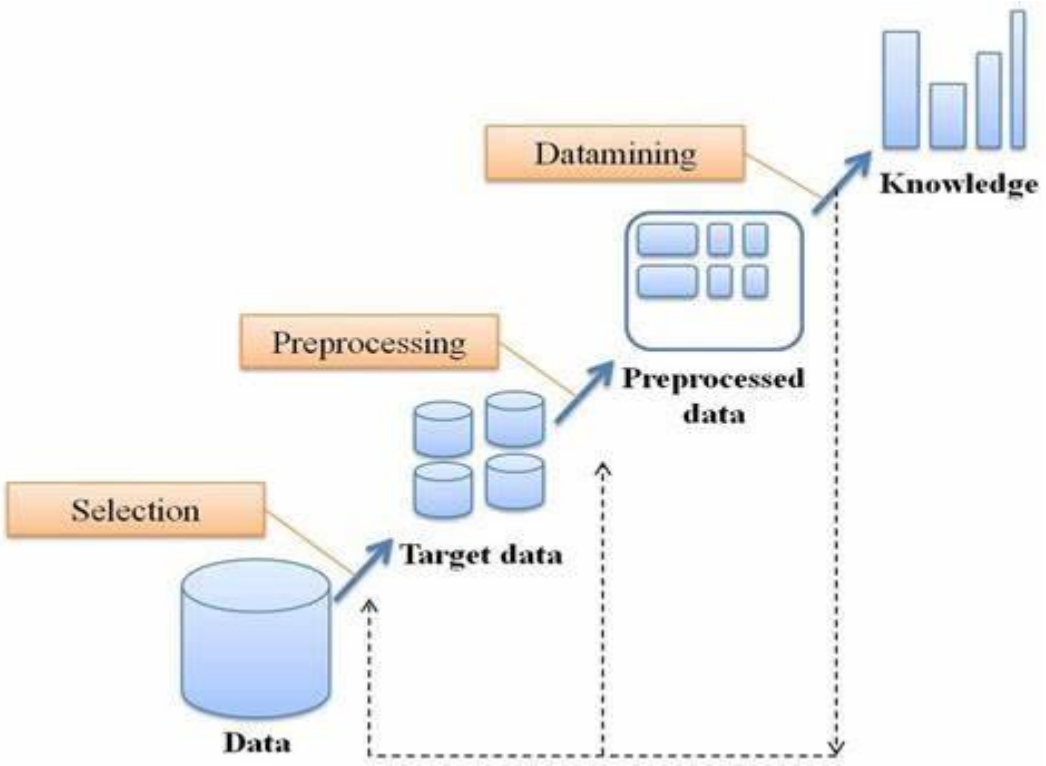

Figure 2: Process of KDD.

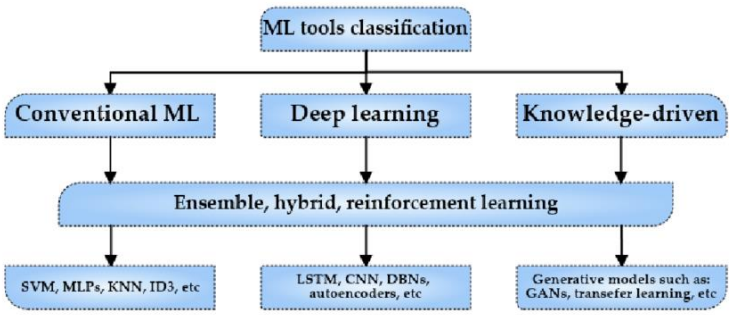

Figure 3: Approaches of DL, ML.

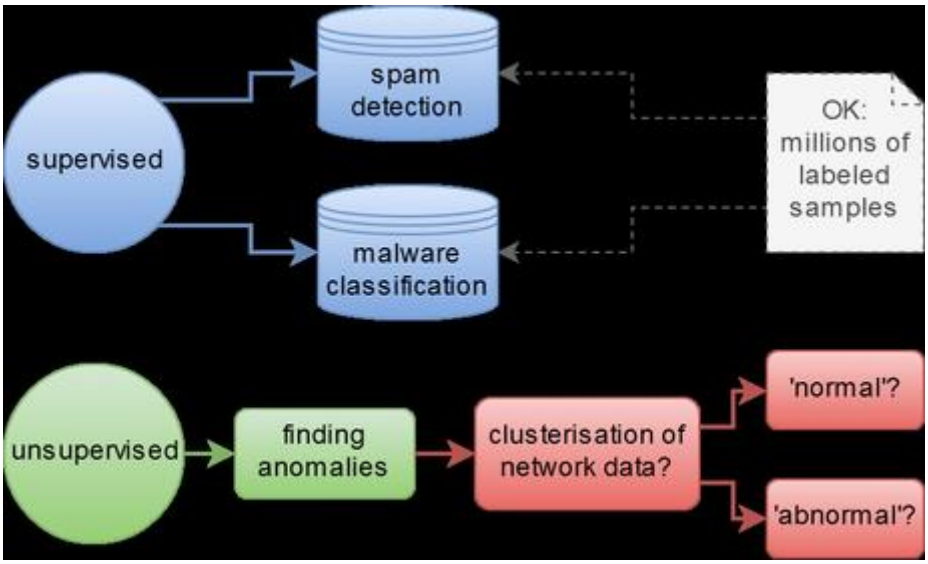

Figure 5: ML in Cyber Security.

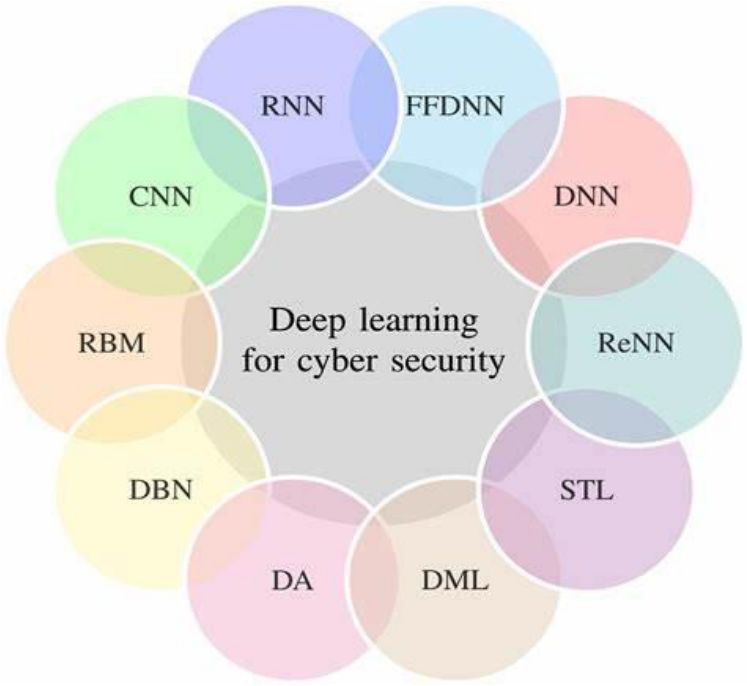

Figure 6: Deep learning for Cyber security.

4. References

1. Kim, M. Park, and D. H. Lee, ‘‘AI-IDS: Application of deep learning to real-time web intrusion detection,’’ IEEE Access, vol. 8, pp. 70245–70261, 2020.
2. Gyusoo Kim and Seulgi Lee, “2014 Payment Research”, Bank of Korea, Vol. 2015, No. 1, Jan. 2015.
3. S. Vishwakarma, V. Sharma, and A. Tiwari, ‘‘An intrusion detection system using KNN-ACO algorithm’’.
4. M. Alkasassbeh and M. Almseidin, ‘‘Machine learning methods for network intrusion detection,’’ 2018, arXiv:1809.02610
5. L. Arnroth and J. Fiddler Dennis, ‘‘Supervised learning techniques: A Comparison of the random forest and the support vector machine,’’ Uppsala Univ., Uppsala, Sweden, 2016.
6. H. Dhillon, ‘‘Building effective network security frameworks using deep transfer learning techniques,’’ M.S. thesis, Dept. Computer. Sci., Western Univ., London, ON, Canada, 2021.
7. https://www.itgovernance.co.uk/cyber-defence-in-depth
8. https://cybersecurityforme.com/cybersecurity-infographics-facts-stats
9. https://www.datto.com/blog/5-amazing-applications-of-deep-learning-in-cybersecurity
10. https://www.sciencedirect.com/science/article/abs/pii/S1574013720304172
11. https://www.isaca.org/resources/infographics/state-of-cybersecurity
12. https://www.researchgate.net