PERFORMANCE EVALUATION FOR DIFFERENT INTRUSION DETECTION SYSTEM ALGORITHMS USING MACHINE LEARNING

Mustafa Nadhim Zarir

A Thesis Introduction In

Meet The Requirements For The Granting Of A Master's Degree In

Faculty Of Electrical And Electronic Engineering / Computer Department

Universiti Tun Hussein Onn Malaysia

December 2018

ACKNOWLEDGEMENT

بسم الله الرحمن الرحيم

In the first place I thank God and praise him for everything. I would like to express my deep thanks and gratitude to my supervisor**: Prof. Madya.Dr.Jiwa Bin Abdullah**, for his valuable guidance and advice. He was very influential in helping me write my research and in the course I chose and gave me many excellent suggestions in writing my research.

Besides my supervisor I want to thank my university. **University Tun Hssein Oon** Malaysia, and I do not forget my professors at the Faculty of Electrical and Electronic Engineering / Computer Department.

In addition, I thank my father, my mother, my brothers, my wife and my children for helping me and my support to complete my studies and get my dream, and I thank everyone who helped me complete my studies and complete my Thesis.

'Praise be to God'

ABSTRACT

Intrusion is a set of operations that decide to compromise the integrity, confidentiality, and convenience of pc resources. This definition ignores the success or failure of those operations, therefore it additionally corresponds to attacks on pc systems. The distinguishing actions that mitigate the operations to compromise the integrity, confidentiality, or convenience of pc resources is termed as intrusion detection. The aim of the Intrusion Detection System (IDS) is to monitor the network system for any sort of attacks. The objectives of this project is to evaluate the performance of various intrusion detection algorithms based on machine learning. The algorithms considered are the Naive Bays Algorithm, Decision Tree Algorithm and Hybrid Algorithm for different datasets. The evaluation of of various intrusion detection algorithms for different datasets is done utilizing a set of performance metrics, including accuracy, precision, central processing unit (CPU) efficiency and execution time. The simulation results showed that the Decision Tree Algorithm achieved the highest precision rate by 95.5% within 42 seconds. For accuracy, it achieved 69.5% and can be considered good, as compared to all other algorithms. Additionally, the Decision Tree Algorithm records 22.5% in CPU utilization. For Naive Byes Algorithm it scored 90% for Precision, 65% for Accuracy and 14% for CPU Utilization. Lastly, for Algorithm, it obtained 91% for Precision76% for Accuracy and 23% for CPU.

ABSTRAK

Pencerobohan adalah satu set operasi yang memutuskan untuk menjejaskan integriti, kerahsiaan, dan kemudahan sumber pc. Takrif ini mengabaikan kejayaan atau kegagalan operasi tersebut, oleh itu ia juga sesuai dengan serangan ke atas sistem pc. Perkara yang membezakan tindakan yang memutuskan untuk menjejaskan integriti, kerahsiaan, atau kemudahan sumber pc dipanggil pengesanan pencerobohan. matlamat produk IDS adalah untuk mematuhi sistem rangkaian untuk sebarang jenis serangan. Objektif projek ini adalah untuk menilai prestasi algoritma pengesanan pencerobohan yang berasas kepada pembelajaran mesin. Algoritma yang dipertimbangkan adalah Algoritma Naive Bays, Algoritma Pokok Keputusan dan Algoritma Hibrid untuk dataset yang berbeza. Penilaian algoritma sistem pengesanan intrusi yang berbeza untuk kumpulan data yang berbeza menggunakan satu set metrik prestasi, termasuk ketepatan, kejituan, , kecekapan unit pemprosesan pusat (CPU) dan masa pelaksanaan. Hasil simulasi menunjukkan perbandingan ketepatan (%) untuk algoritma Tree Decision mencapai kadar Precision tertinggi sebanyak 95.5% melalui 42 saat, Secara Ketepatan, ia mempunyai 69.5% menganggap nilai yang baik dalam semua algoritma yang berbeza untuk Dataset1 dan Dataset2, dari sisi lain ia mencatatkan 22.5% dalam penggunaan CPU. Dalam Algoritma Byes Nave Kedua mendapat tahap tinggi 90% untuk Precision, 65% untuk Ketepatan Nilainya adalah penghampiran sebelumnya, sementara 8second, dalam penggunaan CPU ia mempunyai kadar terendah dalam semua algoritma yang berbeza sebanyak 14%. Dalam Rata-rata Algoritma Algoritma Hibrid tiba di 91% yang merupakan tahap yang baik, melalui 54 kali, Ketepatan dalam Algoritma Hibrid mempunyai nilai tertinggi 76% dalam algoritma yang berbeza secara keseluruhan untuk Dataset1 dan Dataset2, Dalam penggunaan CPU sama dengan perkadaran dengan Algoritma Tree Decision 23%.

TABEL OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

In last year's network security has become very important recently because of the increasing volume and speed of data connection in real time every year. Without strong security systems and through the Internet, hacker can find personal information. That is why many researchers, engineers and developers are working on improving systems. Security in the world to control access to personal information. An incident or breach of confidentiality caused you to access personal resources and access the computer in an unauthorized manner. Infringement or action violates integrity If you allow access, an incident or procedure results in an access violation if legitimate users are not allowed to access resources or services, or reside on a laptop.

In this chapter discusses the background of the Thesis. in addition, the objectives of project, draw back statement and thus the Thesis scopes

1.1 Background of Thesis

In our society, data is changing into more and more addicted to speedy access and interactive process. As this demand will increase, a lot of data is being hold on computers. The proliferation of cheap laptops and computer networks has worsened the matter of unauthorized access and change of state with knowledge. Increasing property not solely provides access to massive and varied sources of information a

lot of quickly than ever before, it conjointly provides Associate in Nursing access path to knowledge from just about any place on the network. With Associate in Nursing accrued understanding of however systems work, intruders became extremely consummate at designation weaknesses in systems, and exploiting them to get such privileges that they'll do something on the system. They conjointly use patterns that are tough to trace and establish. laptop systems are thus not going to stay safe within the nearest future as a result of the intruder's acts. Therefore, we tend to should have measures in situ to sight security breaches, i.e., establish intruders and their strategies of intrusion [1].

When a laptop must perform some task, a programmer's answer is to write down a Trojan horse that performs the task. A Trojan horse may be a piece of code that instructs the pc that actions to require so as to perform the task. the sector of machine learning cares with the higher-level question of a way to construct laptop programs that mechanically learn with expertise. A Trojan horse is claimed to find out from expertise with relation to some category of tasks and performance live P, if its performance at tasks in as measured by improves with expertise. Thus, machine learning algorithms mechanically extract information from computer readable data [2] [3].

In machine learning, pc algorithms (learners) conceive to mechanically distil information from example information predictions regarding novel data within the future and to produce insight into the character of the target ideas [4].

Computer networks are outlined to the variability of sorts and threats that would have an effect on the integrity of the info transmitted, confidentiality of the data and the provision of network services. Network analysis and intrusion detection are countermeasures to thwart bonnets, viruses, spam, worms and the other varieties of malicious attacks on the network. as an instance, the amount of intrusion tries per day in 2003 was within the order of twenty five billion [3].

Researchers mentioned that even one visit to the infected websites allows the offender to sight vulnerabilities of the user's applications and force the transfer a large number of malware binaries. Metrics on web-based malware given within the cited work connected with high confidence that ten percent of the (4.5) million analysed URL are malicious. From Associate in Nursing operational purpose of read, the increasing trend within the range of security controls deployed in networks is

indicating that everyday intruders baffle researchers with new insidious attacks. although previous threats persist, techniques evolved: greed for ill fame, naïve viruses and expressed attacks became less engaging. Economy factors and policy are motivating the interest in concealing and complicated attacks [4]. Figure 1.1 shows the abstract design of learning-enhanced reactive security mechanisms

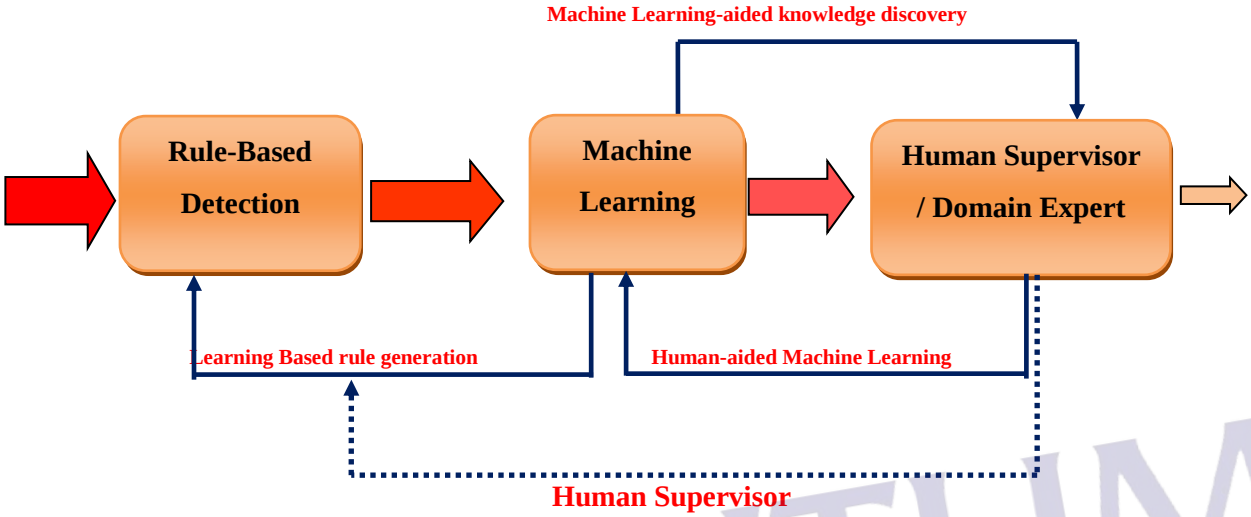

Figure 0.1 Machine Learning Methods for Computer Security [5]

The Internet is accepted the essential tool and one in all the suggested sources of data concerning the present world. net is thought-about collectively of the sphere parts of academic business purpose. Therefore, the information across the web should be secured and therefore the net security is one in all the most important considerations today of all the planet wide security. As net is vulnerable by varied attacks it's terribly essential to style a system to guard those information, yet because the users victimization those knowledge. The Intrusion detection system (IDS) is so AN invention to fulfil that demand work. The network administrations Intrusion detection system so as to stop malicious attacks.

Intrusion Detection (IDS) is the method of monitoring events and detecting unnatural entry that occurs in a system or a group of systems or network computing excessively and analyzing it to detect and identify types of intervention signs. The intrusion detection and intrusion detection system may be a system, a set of systems or a device that examines and tests by automating the way events are monitored and analyzed. With the rise and rise of attacks, many intrusion detection systems are planned within the literature. Despite the diversity and diversity of planned systems

in some respects or in many different aspects, there is a box that measures some of the basic parts that measure the entry process in most of the systems used and planned. Figure 1.2 depicts a really easy generic design of a typical IDS.

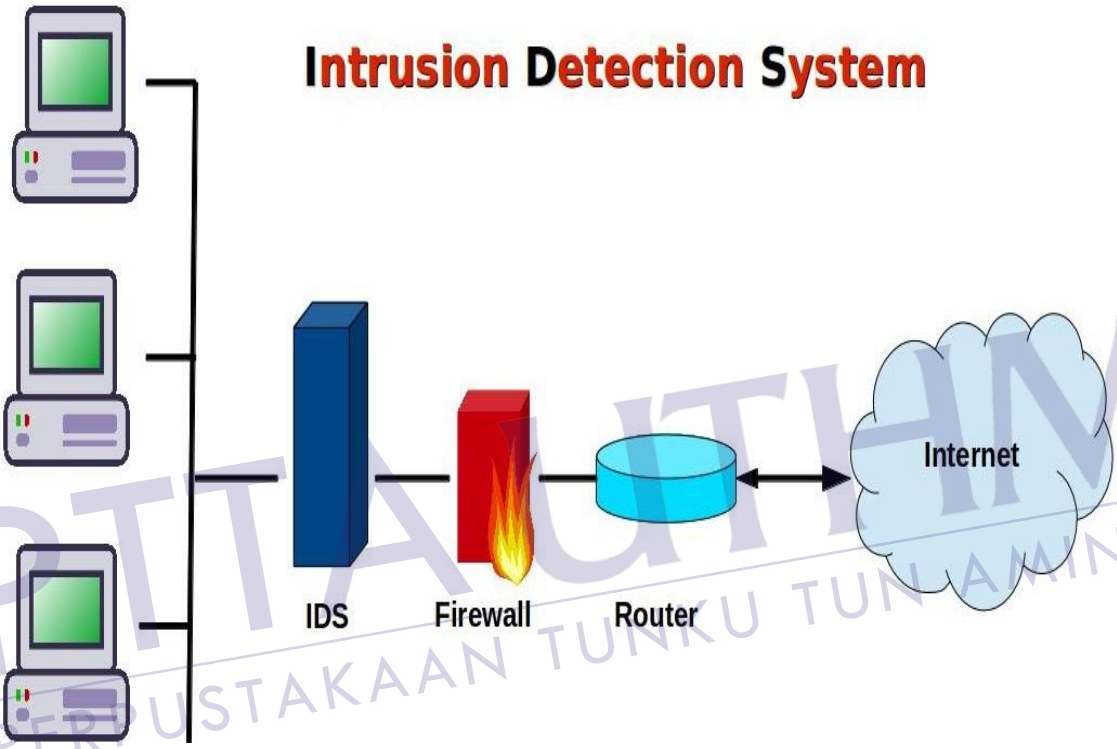

Figure 0.2: A generic architecture [4]

The previous form, the monitoring system is that the system is news and study each entry process is identity protection. They are mostly a single host if they have two or two complete networks. Blogging / Storage Combines information to search, process, and identify events. The process unit is the heart and mind of IDS. Hence all algorithms are square to measure access to find evidence of suspicious behavior. In order to support the efficiency of IDS, an anti-action is often taken by IDS to mitigate the same problem and control it.

1.2 Problem Statement

Keeping networks secure has ne'er been such an indispensable task as nowadays. Threats come back from hardware failures, software system flaws, tentative inquisitor and malicious attacks. With the ever increasing size of networks, many of the current systems employ the signature based intrusion detection system. Such system lacks the power to recognize the new type of threads and requires constant update of the threat signature database to recognize new threats. Hence, due to shortcoming of signature based IDS, machine learning approaches are being used to detect the anomaly and intrusions in the network. The **main problem** is the current algorithms are not benefit for all applications/datasets. Therefore, this project aims to evaluate performance of different intrusion detection system algorithms based on machine learning for computer networks.

1.3 Objectives of the Project

The objectives of this project are:

(I). To identify different of intrusion detection system algorithms for computer networks.

(II). To evaluate the performance of those algorithms for various datasets employing a set of performance metrics through experiment simulations.

(III). To perform assessment analysis of the intrusion detection system algorithms for computer networks.

1.4 Scope of Project

The scopes of this project are:

- Study the Naive Bayes Algorithm, Decision Tree Algorithm and Hybrid Algorithm for different datasets.
- The following datasets are chosen:(i) KDDTrain+.TXT(ii) KDDTrain+_20Percent.TXT(iii) KDDTest+.TXT(iv) KDDTest-21.TXT
- The evaluation of different of intrusion detection system algorithms for different datasets utilizing a set of performance metrics, including accuracy ,Precision ,central processing unit (CPU) efficiency and execution time.
- The NIDS network tool based on JAVA will be used to study and analysis the different of intrusion detection system algorithms.

1.5 The project report outlines

The outlines of the project reported as the follows:

Chapter 1 introduces the project objective, project scopes and problem statements Chapter 2 presents a survey about the major issues in the intrusion detection system and an overview the important Martial's related to the project. The methodology of the project, parameters and simulation tools, which are will be used to implement this project has been explained in chapter 3. I will implement the algorithms with Data Set and analyze their results 4. Finally, the project conclusion and the future works will be done in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

This chapter summarizes the data of the connected previous studies on Intrusion Detection System. additionally, the chapter covers varieties of most Intrusion Detection System, the connected algorithms and therefore the simulation surroundings. These reviews area unit done supported materials from journals, conference continuing and books.

2.1 Overview of Intrusion Detection System (IDS)

Intrusion is any set of procedures that arrange for the safety, confidentiality, and skill of a computer supplier. This definition ignores the success or failure of these actions, and therefore simultaneously corresponds to attacks against a system. The issue of distinguishing between the procedures that result from a breach of safety, theft or the gift of a computer supplier is called intrusion detection. The aim of the IDS product is to monitor the network system for any type of attack. An attack may be referred to with one thing easily such as a program that would modify the user name or it could be an obscene attack involving a series of events stretching across multiple systems. IDSs measure the workbook box through system screens as a result of the fact that they sometimes rely on audit information provided by system records or information collected by inhaling network traffic [4].

But there are detection systems intrusion and hacker that do not work in real time and in real network, either as a result of factors and tools and the nature of analysis and testing they do or as a result of the type of network to be implemented

(analysis of what is going on in the past on a system). The definition of the intrusion detection system does not include AN, which prevents intrusion, although it only abstracts it and processes news. There is a measurement box for some intrusion detection systems that respond once an unauthorized procedure is detected. This reaction sometimes involves stopping the infection, for example, by ending network membership.

IDS is required for a single firewall to protect computer systems. The central components of the intrusion detection engine are: resources that must be protected in the target system excessively, ie user account file systems, system kernel. Models that describe the traditional or legitimate behaviour of those resources; techniques that compare the activities of a particular system with established models, and those who measure the gradient as abnormal or intrusive.

The intrusive analysis process will be separated into four phases as follows:

A. Pre-processing: Once the knowledge is collected from ANS, the information is organized for classification. Pre-processing can facilitate the task of North America to see where information is placed, which is usually a basic format or structured information.

B. Analysis: The analysis phase begins once the pre-processing phase is completed and is applied to any or all records in the information. The log information is compared with the content, as well as log information can be recorded as an entry event or it will be generated.

C. Replies: Once the log information is recorded as an input, the reply will begin. This response contains the alert and negative nesting data.

D. Stimulation: This stage is responsible for the health of parasitism.

2.2 Types of Intrusion Detection Systems(IDS)

2.2.1 Intrusion detection systems can be classified into different categories depending on how they work [12].

2.2.2 Network/Host Based IDS

Based on the resource collection and storage unit, we discover that we have two IDS systems. Network-based IDS collects and stores information used directly from the network being monitored and controlled, in the packet mode window. For this reason, any NIDS is actually automatically running. Most NIDS are independent of the operating system, and for this reason, simple to deploy and use. It provides the highest level of security and privacy against DoS attacks. However, there is a problem in this type where this type of IDS cannot examine the protocols for intruders or content if the network traffic to monitor is encrypted. Intrusion detection becomes more robust on modern switching networks, as packages are not available for NIDS.

Under a similar and similar level, and the second type is Host-Based IDS (HIDS). HIDS collects and finds information from the host, which is protected, within the system in the format of the operating system log file to be monitored, system calls, and C.P.U. Use NT event logs, application level logs, etc. Here we have a problem where these systems are not effective by encrypted traffic or switched networks. However, the HIDS depend on the operating system of the network to be controlled, and therefore need some before going out before execution .These systems are extremely useful and economical in attacks that bypass the buffer of detective work [12].

2.2.3 Misuse/Anomaly Based IDS

Another normal for IDS classification is from the treatment / detection purpose of read. supported detection technology, there are 2 sorts of IDS. supported misuse, conjointly called signature-based, IDS maintains a info of signatures of well-known attacks. once the info is received from the audit unit, it matches the info against the info and if any match is found if found, the alarm is triggered. For abuse-based detection, making / linguistic communication signatures could be a tough task, and most analysis focuses on this issue. it's clear that this kind of IDS is unable to sight zero-day attacks, as a result of the signatures of those attacks don't seem to be offered in its info. however the most effective issue regarding this kind of symbol is that the warning rate is simply too low. Most industrial IDSs of this class [12].

The second kind of category is IDS-based anomalies, conjointly called behaviour-based systems. rather than maintaining signatures of well-known attacks, these systems recognize the natural behaviour of the entity and area unit monitored, within the sense that they maintain signatures of traditional behaviour. Any deviation from traditional behaviour is suspicious and therefore the alarm is about. These systems work on the belief that any abnormal behaviour or activity is considerably completely different from traditional behaviour. By definition, these systems area unit capable of detective work zero-day attacks. however these systems suffer from a high rate of false alarms as a result of any deviation from traditional activity might not be interference. Therefore, reducing false alarms is that the focus of analysis at intervals these systems.

There also are another criteria for classifying IDS into completely different classes. as an example, Depending on the response, IDS are often passive or active. [9].

2.3 What Are The Machine Learning (M L) Techniques

Machine learning methods measure completely different methods of automated learning that you use and how. There is a square measuring many of the ways in which the learning is done under various techniques. Because of the excessive availability of digital files and the desire of the accompanying structure, the automatic classification (or classification) of texts into predetermined categories

has raised interest over the last 10 years. In the analysis society, most of the techniques of this defect rely on automated learning techniques: the method of universal induction automatically adopts workbooks by learning the properties of layers from a set of pre-categorized documents. The benefits of this technique on information engineering methods (industrial definitions of works by field experts) a highly effective square measurement, will greatly save the work of the specialist, and can be transplanted directly to completely different fields. This research discusses one of the methods of classification of data belonging to the model of automatic learning. There are some problems associated with 3 completely different issues, notably illustration, classification building, and classification analysis [12]. This type of learning is also referred to as learning from examples. In supervised learning, the system of psychological traits must recognize the ideas or functions that describe the descriptions of the model. In particular, the system provides a set of examples. The outcome of the objective applies further to each of these examples. The system should note that the model diagram supports performance output. For analysis functions, the information set (training set) is used to create the model, while the remainder of the information is used to judge the designer model (test set). The establishment of two educational tasks, especially classification and analysis, in the direction of automatic learning. The classification involves the development of diagnostic models with distinct ranges of functions, while the analysis involves the development of models with varying infinitely different functions. The most common ways to learn box direction learning box are as follows:

- Concept learning. psychological feature system provides samples of happiness (positive Example) or doesn't belong to (counterexample) construct (class). Then, the system is invoked to come up with a generalized description of the construct in order that future cases will be set supported this description.
- Classification or call tree induction. Classification or call tree induction ways very hip and accustomed approximate distinct target functions. These ways build a tree structure that diagrammatically represents coaching information. the most advantage the choice tree is that they're straightforward to clarify. call trees can even be pictured as "if-then "rule.

- Rules learning. Rule learning includes the induction of "if - then" rules, referred to as classification rules. Classification rules square measure accustomed approximate the distinct objective perform.
- Instance-based learning. during this learning, information is hold on in its original format. once the system is termed to make a decision on a replacement case and it examines the link between the new cases every hold on example. Since then, this type of learning is additionally referred to as lazy learning. the training method was delayed till new cases emerged.
- Bayesian learning. This learning is predicated on theorem and contains some ways Use chance. Existing information will be incorporated within the variety of initial chance.
- Linear regression. regression toward the mean may be a technique of describing the target perform linearly combination of another variables. The scope of the target perform should be continuous interval.
- Neural Networks. Neural networks will be used for classification and regression. its
- Functions supported biological patterns and varied programs that simulate the human brain Activities square measure used.

The KDD information set contains a smart understanding of varied intrusions, and it's additionally wide accustomed take a look at and value many areas of intrusion detection algorithms. The KDD data set was first published in 1999 by the University of California at its Institute of Technology (Massachusetts Institute of Technology) At the Lincoln Laboratory, [13]. Contains 48843413 cases with 41 attributes used. In this work, the KDD data set was imported to SQL Server 2008 to perform and calculate various statistical measures such as the distribution of instances, methods and types of attacks and the time they occur to reduce them.

2.3.1 What are the uses of the learning machine at present

Machine learning(M L) is employed to try to many various fields in today's society. it's additionally wont to calculate projected functions as several things that it may be used on the device is a set of applications that contain a number of programs used.

- Identification and detection of frauds.
- Filter and sort web search.
- Ads and programs in real time.
- Analyze and arrange text.
- Facebook and many Social Media programs use it for news stories and are useful in the chat column.
- Identify and discover on many styles and images.
- Email filtering and many spam.
- In Devices used in health care and in the medical field
- Monitor and detect intrusion and attempt to enter by hacker to the network.

As may be seen from the list on top of there's a large space that machine learning is employed and it keeps on and gets larger as machines and technologies have gotten higher and quicker. Machine Learning plays a key role in several scientific disciplines and its applications are a part of our lifestyle. it's used as an example to filter spam email, for weather prediction, in diagnosis, product recommendation, face detection, fraud detection, etc [14].

Machine Learning (ML) studies the matter of learning, which may be outlined because the downside of feat information through expertise. This method generally involves perceptive a development and constructing a hypothesis thereon development which will permit one to create predictions or, a lot of generally, to require rational actions. For computers, the expertise or the development to find out is given by the information, therefore we are able to outline millilitre because the method of extracting knowledge from information. Machine learning is closely associated with the fields of Statistics, pattern recognition and data processing. At

constant time, it emerges as a subfield of technology and offers special attention to the recursive a part of the information extraction method. In summary, the main focus of millilitre is on algorithms that ar ready to learn mechanically the patterns hidden within the information..

2.3.2 Learning under supervision

Learning can be supervised by a machine learning method, supervised by a word that says there is a variety of training assistance within the approach that works anyway. There may be a police investigation approach that is assisted by the engine by marking the information that is used to use it. This may be a good use if a user tries to see the most common, or least common, or least common colour for our cars using a traffic camera. They can then label and categorize the different car colors and match those colors to the cars with the vehicles on the previously stored traffic camera. This could be just one example illustrating the process of using supervised learning. Supervised education makes detection or higher knowledge easier and simpler. There are square management expressions, for example, a car may be a predefined colour, or traffic may be a method of attack. If the information you attend is marked for use, the automated learning technique used here is supervised learning.

We also have many learning algorithms under supervision such as artificial neural network, theoretical statistics, Gaussian regression method, lazy learning, neural neighbor formula, vector support machine, hidden André model off, theoretical networks, call trees (C4.5, ID3, (Randomized), K-nearest Neighboring, Boosting, Ensembles classifiers (Bagging, Boosting), Linear Classifiers (Logistic Revression, Fisher Linear discriminate, Naive Bayes Categorization, Perception, SVM) Supervising algorithms.

2.3.3 Learning without supervision

Uncontrolled learning is that the opposite of supervised learning, rather than simply examining one of the signs, completed in supervision, is in the overall picture. However, in cases of unattended educational information is not registered. A great

REFERENCES

1. Munoz, A. (2014). Machine Learning and Optimization. URL: https://www. cims. nyu. edu/~ munoz/files/ml_optimization. Pdf [accessed 2016-03-02][WebCite Cache ID 6fiLfZvnG].
2. Haq, N. F., Onik, A. R., Avishek, M., Hridoy, K., Rafni, M., Shah, F. M., &Farid, D. M. (2015). Application of machine learning approaches in intrusion detection system: a survey. International Journal of Advanced Research in Artificial Intelligence.
3. Estrada, V. C., & Nakao, A. (2010, January). A survey on the use of traffic traces to battle internet threats. In Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on (pp. 601-604). IEEE.
4. Estrada, V. D. C. (2017). Analysis of Anomalies in the Internet Traffic Observed at the Campus Network Gateway. arXiv preprint arXiv: 1706.03206.
5. Joseph, A. D., Laskov, P., Roli, F., Tygar, J. D., & Nelson, B. (2013). Machine learning methods for computer security (Dagstuhl Perspectives Workshop 12371). In Dagstuhl Manifestos (Vol. 3, No. 1). SchlossDagstuhl-Leibniz-ZentrumfuerInformatik.
6. Aleem, S., Capretz, L. F., & Ahmed, F. (2015). Benchmarking machine learning techniques for software defect detection. Int. J. Softw. Eng. Appl, 6(3).
7. Pundir, P., Gomanse, V., &Krishnamacharya, N. (2013). Classification and Prediction techniques using Machine Learning for Anomaly Detection. International Journal of Engineering Research and Applications (IJERA).
8. Bengio, Y. (2009). Learning deep architectures for AI. Foundations and trends® in Machine Learning, 2(1), 1-127.
9. Dal Pozzolo, A., &Bontempi, G. (2015). Adaptive machine learning for credit card fraud detection.
10. Z. Zhi-Hua, L. Hang, Y. Qiang, "Advances is Knowledge Discovery and Data Mining," 11th Pacific-Asia Conference, PAKDD, Springer, China, Vol.4426, 2007.

11. L. Wenke, J. Salvatore, W. Mok, “A Data Mining Framework for Building Intrusion Detection Models, “IEEE Symposium on Security and Privacy, pp. 120-132, 1999.
12. M. M. Gaber,” Advances in Data Stream Mining”, John Wiley and Sons, Vol. 2, No. 1, pp. 79-85, 2012.
13. A. N. Huy, D. Choi, “ Application of Data Mining to Network Intrusion Detection: Classifier Selection Model”, Asia-pacific Network Operation and Management Symposium (APNOMS), Springer-Verlag Berlin, Heidelberg, pp. 399-406, 2008.
14. M. Dewan, H. Nouria, B. Emna, Z. Mohammed, M. Chowdhury, “ Attacks Classification in Adaptive Intrusion Detection using Decision Tree”, World Academy of Since, Engineering and Technology, No.63, pp.27-44, 2010a.
15. Sommer, R., &Paxson, V. (2010, May). Outside the closed world: On using machine learning for network intrusion detection. In Security and Privacy (SP), 2010 IEEE Symposium on (pp. 305-316). IEEE.
16. Shon, T., & Moon, J. (2007). A hybrid machine learning approach to network anomaly detection. Information Sciences, 177(18), 3799-3821.
17. Shon, T., Kim, Y., Lee, C., & Moon, J. (2005, June). A machine learning framework for network anomaly detection using SVM and GA. I Information Assurance Workshop, 2005. IAW'05. Proceedings from the Sixth Annual IEEE SMC (pp. 176-183). IEEE.
18. Gandhi, G. M., & Srivatsa, S. K. (2010). Adaptive machine learning algorithm (AMLA) using J48 classifier for an NIDS environment. Advances in Computational Sciences and Technology, 3(3), 291-304.
19. Yu, Z., & Tsai, J. J. (2008, June). A framework of machine learning based intrusion detection for wireless sensor networks. In Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC'08. IEEE International Conference on (pp. 272-279). IEEE.
20. Goyal, A., & Kumar, C. (2008). GA-NIDS: a genetic algorithm based network intrusion detection system. Northwestern university.
21. Song, J., Takakura, H., Okabe, Y., Eto, M., Inoue, D., & Nakao, K. (2011, April). Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation. In Proceedings of the First Workshop on

Building Analysis Datasets and Gathering Experience Returns for Security (pp. 29-36). ACM.

22. Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.
23. Tzanis, G., Katakis, I., Partalas, I., &Vlahavas, I. (2006, July). Modern applications of machine learning. In Proceedings of the 1st Annual SEERC Doctoral Student Conference–DSC (Vol. 1, No. 1, pp. 1-10).
24. L. Wenke, J. Salvatore, W. Mok, "A Data Mining Framework for BuildingIntrusion Detection Models, "IEEE Symposium on Security and Privacy, pp. 120-132, 1999.
25. M. Tatsuya, I. Masayuki, N. Ayahiko, K. Osmu, "Extension of Decision TreeAlgorithm for Stream Data Mining Using Real Data," 5th InternationalWorkshop on Computational Intelligence and Applications (IWCIA). IEEESystems, pp. 208-212, 2009.
26. B. Daniel, J. Couto, S. Jajodia, N. Wu, "ADAM: A Test bed for Exploring theuse of Data Mining in Intrusion Detection", SIGMOD, Vol. 30, No. 4, 2001.
27. H. Sang-Jun, C. Sung-Bae, "Combining Multiple Host-Based Detectors usingDecision Tree," 16th Australian Conference on Artificial Intelligence, Springer,Vol. 2903, pp.208-220, 2003.
28. S. Ahmed, "Intrusion Detection System using Data Mining", Applied ScienceDepartment, Master Thesis, University of Technology, 2006.
29. B. Yacine, C. Frederic, "Neural Networks vs. Decision Tree for IntrusionDetection," IEEE/IST Workshop on Monitoring Attack Detection and Mitigation(MonAM), 2006.
30. P. Mrutyunjaya, R. P. Manas," Network Intrusion Detection using NaïveBayes", International Journal of Wei-computer Science and Network Security(IJCSNS), Vol. 7, No.12, pp. 455-464, December 2007.
31. A. N. Huy, D. Choi, " Application of Data Mining to Network IntrusionDetection: Classifier Selection Model", Asia-pacific Network Operation andManagement Symposium (APNOMS), Springer-Verlag Berlin, Heidelberg, pp.399-406, 2008.
32. M. Dewan, H. Nouria, B. Emna, Z. Mohammed, M. Chowdhury, " AttacksClassification in Adaptive Intrusion Detection using Decision Tree",

WorldAcademy of Since, Engineering and Technology, No.63, pp.27-44, 2010a.

33. R. Hanumantha, G. Srinivas, D. Ankam, K. Vikas, " Implementation ofAnomaly Detection Technique using Machine Learning Algorithms",International Journal of Computer Science and Telecommunications (IJCST),Vol. 2, Issue 3, pp. 25-31,2011.
34. R. Shan mugavadiru, N. Nagarajan , " Network Intrusion Detection System usingFuzzy Logic," Indian Journal of Computer Science and Engineering (IJCSE),Vol. 2, No. 1, pp. 101-111,2011.
35. G. Radhika, S. Anjali, C. J. Ramesh, "Parallel Misuse and Anomaly Detection Model," International Journal of Network Security, Vol.14, No.4, pp. 211-225,July 2012.