# PREDICTION OF NETWORK ATTACKS USING SUPERVISED MACHINE LEARNING TECHNIQUE

[1]ABBURI SAMPATH KUMAR, [2]TALLURI S S SAI KRISHNA, [3]THELLAGORLA RAMGOPI, [4]DUDEKULA MABU RAMJAN, [5]DR. C. RAJABHUSHANAM

[1,2,3,4]STUDENTS, [5]PROFESSOR
BHARATH INSTITUTE OF HIGHER EDUCATION AND RESEARCH

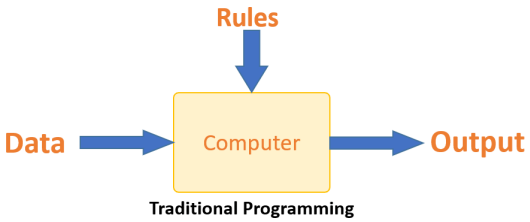
*Abstract*- **With the development of wireless communications at the Internet, there are numerous security threats. An IntrusionDetection System (IDS) helps stumble onattacks on a system and discover intruders. Previously, numerous gadget learning strategies (ML) were used in IDS strategies which have attempted to enhance intruderdetection consequences and improve the accuracy of IDS. This article presents an approach to imposing an IDS the usage of Principal Component Analysis (PCA) and a random forest type set of rules. Where PCA will assist to arrange the information withthe aid of lowering the dimensionality of thedata and Random Forests will assist inside the type. The outcomes obtained show that the proposed approach plays extra efficiently in terms of accuracy compared to other strategies, which include SVM, Naïve Bayes and Decision Tree. The results received via the proposed approach have values for the length (min) of 3.24 mins, accuracy (%) of 96.78% and accuracy (%) of 0.21%.**

**Keywords: Machine Learning (ML), classification method, python, Prediction of Accuracy result.**

## INTRODUCTION

A device getting to know system is a laptop algorithm that may study through example thru self-improvement without being explicitly marked through a programmer. Machine learning is a part of artificial intelligence that mixes facts with statistical gear to expect outcomes that can generate actionable insights.

The department comes with the concept that a machine can analyze from information (i.E., by way of example) about itself to supply particular outcomes. Machinestudying is closely related to statistics mining and Bayesian predictive modeling.




The device takes enter and uses an set of rules to provide solutions.

The education device have to make an ordinary advice. For those with a Netflixaccount, all film or series hints are based at the user's historical records. Companies are the usage of non-associative gaining knowledge of techniques to improve person enjoy with personalized hints.

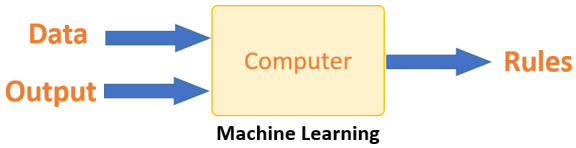
Machine studying is also used in diverse duties such as fraud detection, preventivepreservation, portfolio optimization, work automation, and so forth.

### Machine Learning vs. Traditional Programming

Traditional programming differs appreciablyfrom machine studying. In traditionalsoftware, the programmer codes all of the regulations in session with an professional inthe enterprise for which the programmer is developing. Both regulations are based totally on a logical foundation; the system will output the subsequent common-senseoperator. As the machine becomes greater complex, greater policies should be written. It can quick turn out to beunstable. Traditional programming differs considerably from gadget mastering. In traditional software, the programmer codes all the rules in consultation with an expert inthe enterprise for which the programmer is growing. Both regulations are based totally on a logical basis; the device will output the following common-sense operator. As themachine turns into extra complicated, greater guidelines must be written. It canspeedy become volatile.

Traditional Programming

Machine gaining knowledge of is supposed to clear up this trouble. The machine learns how the enter and output are associated and writes the rule. Programmers ought not towrite new rules whenever they enter newrecords. Algorithms adapt in response to new records and reports to enhance time efficiency.

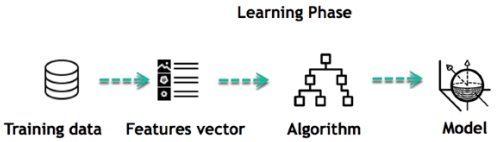



Machine Learning

**How does Machine Learning Work?**

The getting to know apparatus is the brain where all mastering takes area. A systemlearns in a human-like manner. They analyze from revel in. The more we realize, the easier it's far to predict. Similarly, when confronted with an unknown state of affairs, the probability of fulfillment is decrease than in a recognized scenario. Machines research in the equal way. The machine sees the sample to make an correct prediction.When we supply a comparable example to the machine, it is able to calculate the end result. But, like a man or women, if he feedsa pattern he hasn't seen earlier than, it isdifficult for a device to predict.

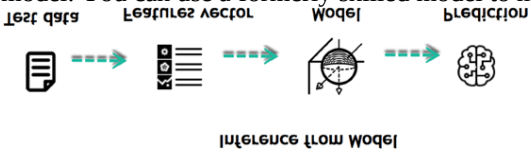
The primary intention of device mastering islearning and inference. First, the machinelearns by way of locating patterns. This discovery changed into given thank you.One of the most important responsibilitiesfor a facts scientist is to carefully pick outwhich statistics engine to offer. The listingof attributes used to remedy a trouble isknown as a characteristic vector. You canconsider a function vector as a subset ofinformation this is used to remedy a trouble. A prodigious gadget uses algorithms to simplify things and turn this discovery into aversion. Therefore, the education segmentdescribes the information and is usually described in the version.




For example, the engine looks for a relationship among a person's profits and thelikelihood of going to a present day restaurant. It seems that a machine that reveals the connection between salaries and going out to an costly eating place: it is a model

**Inferring**

Once the model is built, you may check howpowerful it's miles on formerly unseeninformation. The new information is transformed into a characteristic vector,handed to the model, and made a prediction. All of this is a lovely part of system studying. There is no want to update therules or maintain the model. You can use a formerly skilled model to make guesses on new records.



Lifetime learning applications are simple and can be summarized within the followingpoints:

1. Define the problem
2. Address given
3. Visualize the information
4. Learning set of rules
5. Test algorithm
6. Collect critiques
7. Refine the set of rules
8. Cycle four-7 till pleasant effects.9 Use the model for prediction

As the algorithm learns to attract correct conclusions, it applies the expertise to new datasets.

**LITERATURE SURVEY**

**1)** A Proposed Wireless Intrusion Detection Prevent ion and Attack System

**AUTHORS:** JafarAbo Nada; Mohammad Rasmi Al-Mosa

This electronic mail report is a "living" template and already defines the elements ofyour article [title, text, headings, etc.] in the fashion sheet. With the rapid deployment of wireless networks, the idea of networksecurity has faced many dangers. And consequently, have to offer security answers.Classical methods of protective networks from assaults are not suitable. For instance, an intrusion detection machine that works onstressed out networks is rendered useless on wi-fi networks. Wireless era has opened a brand-new area for network customers. Withits ease of use and customization, this approach has grow to be famous and is converting rapidly. But the concern of the earth, and the primary fear. The purpose for that is because of the decoration. With developing difficulty, you want to think about a safety solution. This article proposesa new intrusion and assault preventionsystem for improving wi-fi networks.Therefore, the object will speak theimprovement of a wi-fi intrusion detection system, that is a wi-fi intrusion and attack prevention device "WIDPAS". It is primarily based on three most important obligations: tracking, analysis and safety. With it, it video display units denial-of- service or fake network assaults, then captures the assault and identifies the attacker, in addition to protects network customers.

**2)** Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm

**AUTHORS:** Kinam Park; Youngrok Song; Yun-Gyung Cheong
In this text, we present the consequences of our experiments to evaluate the effectiveness of detecting distinct styles of assaults (e.G. IDS, malware and shell). We examine the popularity performance with the aid of applying the Random Forest set ofrules to numerous information generated from the Kyoto 2006+ dataset, that is the cutting-edge network document records amassed for the development of intrusiondetection systems. We conclude withdiscussions and future research tasks.

**3)** On the Selection of Decision Trees inRandom Forests
**AUTHORS:** S. Bernard, L. Heutte and S. Adam
In this paper, we present a observe of a own family of random woodland (RF) matching methods. In the "classical" RF induction method, a set variety of decision trees are triggered to form an ensemble. This kind of set of rules has two primary dangers: (i) the wide variety of trees is constant a priori (ii) the interpretation and analysis opportunities which might be lost through the decision tree classifiers because of the precept of randomization. This manner, by which trees are introduced with out consensus, does now not guarantee that every one those timberwill cooperate successfully within the identical plan. This thought increasesquestions: are there any decision timber in RF that result in poor overall performance ofelections? If so, is it feasible to form a more correct committee through doing away with the low performance decision bushes? The solution to these questions is solved as asorting question. Thus, we show that ideal selection timber may be acquired even the usage of a suboptimal classifier selection method. This proves that the "classical" RF induction manner, whereby random bushes are randomly introduced to the ensemble, is not the nice technique for growing correct RF classifiers. We also gift an hobby in RF development, by means of including timber in a manner that is greater based than inconventional "classical" RF induction algorithms.

**4)** Intrusion Detection using Random ForestsClassifier with SMOTE and Feature Reduction
**AUTHORS:** A.Tesfahun,D.Lalitha Bhaskari
Intrusion detection systems (IDS) have turn out to be an essential part of laptop and network safety. The NSL-KDD intrusion detection dataset, that is an prolonged version of the KDDCUP'ninety nine dataset,changed into used as an experimental devicein this newsletter. Due to the inherentcharacteristics of intrusion detection, there may be nevertheless a big imbalance among training in the NSL-KDD dataset, which makes it difficult for system gaining knowledge of inside the field of intrusiondetection. When considering rank inequality,this article applies the Synthetic MinoritySampling (SMOTE) technique to thetraining dataset. An records-based feature choice method is offered for the design ofthe NSL-KDD function-decreased database. Random forests are used as a classifier for intrusion detection purposes. The empirical effects display that the Random Forestclassifier with SMOTE and feature choice based on information acquisition providesthe first-class performance in developing an IDS that is green and powerful for network intrusion detection.

**5)** The Impact of PCA-Scale ImprovingGRU Performance for Intrusion Detection
**AUTHOR:** Let.-T.-H., Kang H. And KimH.
A device or software program package that video display units network or structures for malicious hobby is an intrusion detection machine (IDS). Conventional IDSs do nolonger hit upon sophisticated cyber attacks inclusive of low frequency DoS attacks or unknown assaults. Over the years, devicegaining knowledge of has generated moreand more interest in overcoming these limitations. In this text, we proposed a new approach to enhance Gated Recurrent Unit (GRU) intrusion detection by way of incorporating the proposed PCA-Scale with versions, inclusive of PCA-Standardized andPCA-MinMax, into the GRU layer. Both complementary methods explicitly practice maps of learned object functions, movingwithin the direction of maximum variance with advantageous covariance. This technique may be applied to the GRU modelwith little additional computational price. We present experimental effects on two actual tables, including KDD Cup 99 andNSL-KDD, demonstrating that the GRU model trained with the PCA-Scaled technique makes exceptional progress.

**SYSTEM REQUIREMENTS: HARDWARE REQUIREMENTS:**

- System: Pentium IV 2.4GHz.
- Hard Disk : 40 GB
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb.
-

**SOFTWARE REQUIREMENTS:**

| | | |
|---|---|---|
| Operating system | : | Windows 7. |
| Coding Language | : | Python |
| Database | : | MYSQL |

**SYSTEM ANALYSIS**
**EXISTING SYSTEM:**
Iftikhar Ahmad et al, investigated variousmachine mastering algorithms for intrusion detection system. They in comparison numerous methods which include SVM, Extreme Learning Machine and Random Forest. The authors of the effects state that the Extreme system gaining knowledge of approach plays tons higher as compared to different algorithms.

B. Riyaz et al., labored right here onimproving the high-quality of the facts sets to provide them with an intrusion detection gadget. Although rules had been used from the feature choice technique to enhance the information set. They used the KDD dataset and validated a dynamic boom in IDS consequences.

**DISADVANTAGES OF EXISTINGSYSTEM:**

Systems jogging over the Internet are vulnerable to diverse malicious activities. The major trouble seen in this regard is the intrusion into the data system.

The present results indicate that some enhancements can be made in phrases ofaccuracy, detection fee and false positive charge. Some different strategies can update preceding techniques including SVM and Naïve Bayes. Also, the take a look at says that the dataset can be stepped forward through the use of sure methods in it. Increase the best of enter into the proposed device.

**PROPOSED SYSTEM:**

The intrusion detection system works to enhance the gadget being affected. This detection system can do the trick. The proposed system tries to eliminate problems related to previous operations. The proposedsystem consists of methods: primary aspect analysis and the random woodlandtechnique.

Principal factor evaluation is used to reduce the dimensionality of the dataset; with this approach the best of the dataset can be advanced, as the dataset can contain the perfect attributes. After this, a randombounce set of rules could be implemented to hit upon intruders, which provides each speed and fake effective rate in a better way compared to SVM.

**ADVANTAGES OF PROPOSEDSYSTEM:**

- The mistakes price discovered in our proposed approach may be very low at 0.21%.
- In addition, the accuracy of theresulting algorithms is lots higher than the preceding one.
- In addition, the execution time is lessthan different algorithms.
-

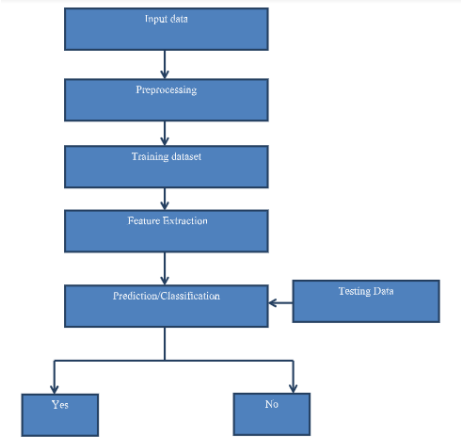
**OBJECTIVES**

1. Input layout is the process of remodeling an input description right into a pc device. This approach is critical to avoid mistakes within the facts access manner and to factor the proper direction to the management to get an appropriate records from theautomatic system.
2. This is carried out with the aid of developing appropriate records access shelves to technique massive amounts of facts. The purpose of the input approach is to simplify statistics access and put off errors. These facts access screen is designedso that all statistics operations may be completed. It also presents a method to viewinformation.
3. When data is entered, it is checked for validity. Data may be entered via screens. Appropriate instructions are furnished as wished, so that the person will not be in an instantaneous country. So the cause of the enter layout is to create an input layout that is simple to observe.

**DATA FLOW DIAGRAM:**

1. A DFD is also known as a bubble chart. Itis a easy graphical formalism that can beused to symbolize a gadget in terms of inputs to the machine, the various processes done on that records, and the outputs generated by it.
2. Data glide diagram (DFD) is one of the major modeling gear. It is used to version parts of the machine. These additives are the system strategies, the information utilized bythe method, the outside item that corresponds to the system, and the data flows inside the machine.
3. The DFD suggests how informationactions thru the device and how it's miles changed via a chain of changes. It is a graphical technique that depicts the float of facts and the adjustments which are applied as information moves from enter to output.

Four. A DFD is also known as a bubblechart. A DFD may be used to symbolize a device at any stage of abstraction. A DFD may be divided into layers that constitute incremental statistics flow and character operations.




**ULM DIAGRAMS**

notation to design software programprojects.

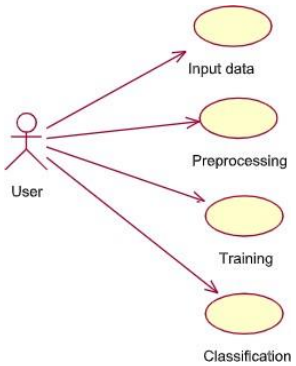
**GOALS:**

The essential desires of UML improvementare as follows:

1. Provide users with a ready-to-use expressive language of visual layout in orderthat significant examples may be evolvedand shared.
2. Provide enlargement and specialization ofengineering equipment to extend core ideas.
   UML stands for Code of Canon Law. UML is a preferred cause modeling language for object-orientated software program improvement. The flag is controlled and created by way of the object control institution.
   UML is meant to emerge as a commonplace language for growing object-orientated laptop software fashions. In its contemporary form, UML has two most important components: the metamodel and the notation. Certain techniques or varieties of strategies may also be brought in the future; or to the UML.
   The Unified Modeling Language is a standard language for expressing,visualizing, constructing, and documenting the structure of software program systems, as well as for modeling commercial enterprise and different non-softwareprogram systems.
   UML Sets engineering best practices which have demonstrated to be effective in modeling big and complex systems.
   UML is an important a part of item- orientated software program improvement and the software program development technique. UML particularly uses graphical
   Be independent from specificprogramming languages and the improvement method.
3. Provide a formal basis for expertise language formation.
4. Strengthen the increase of the marketplacefor OOP tools.
5. Support higher-degree improvementstandards, inclusive of collaboration,frameworks, models, and components.
6. Complete with the nice capabilities.

**USE CASE DIAGRAM:**

The Unified Modeling Language (UML) usecase diagram is a form of human diagram defined and constructed from use case evaluation. The purpose is to offer a graphical overview of the functionality ofthe device in phrases of actors, their goals (represented as use cases), and any dependencies among person instances. The primary use case of a diagram is to expose which system functions are completed forwhich actor. You can describe the roles of the actors inside the device.

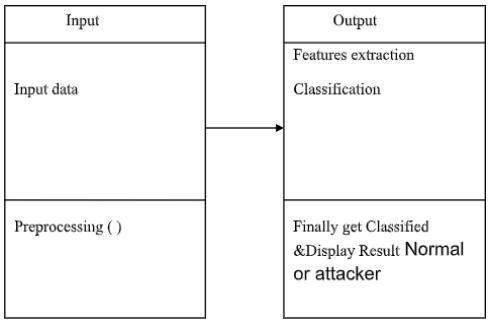


And timing diagrams.

**CLASS DIAGRAM:**

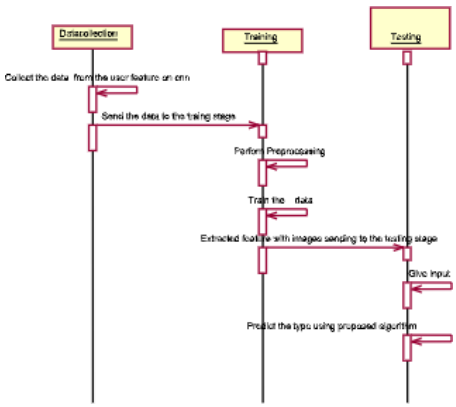
In software program engineering, a Unified Modeling Language (UML) magnificence diagram is a form of static structural diagram that describes the structure of a machine by means of showing the machine'sinstructions, their attributes, operations (or strategies), and relationships among classes.

.It explains what type of statistics it incorporates.



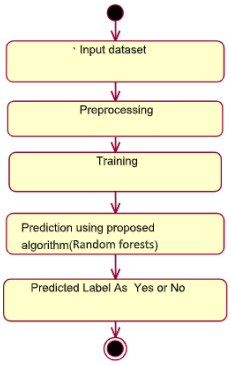

**SEQUENCE DIAGRAM:**

A Unified Modeling Language (UML) sequence diagram is a kind of interplay diagram that indicates how approaches have interaction with each other and in what order. This submit is a series of posts. Sequence diagrams are every so often known as event diagrams, occasion scripts,



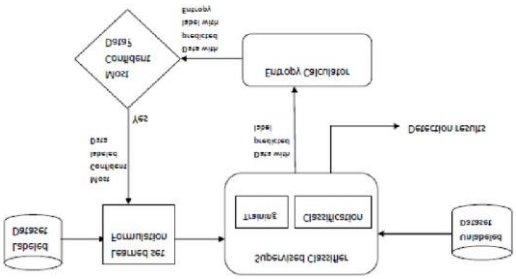

**ACTIVITY DIAGRAM:**

Activity charts are a graphical representationof step-by way of-step and operating sports with help for choice, iteration and concurrency. In a completely unique modeling language, an activity diagram can be used to describe the operations and step-through-step workflow of components in a gadget. The movement diagram suggests theoverall glide of control.




**SYSTEM DESIGN**

**SYSTEM ARCHITECTURE:**



**REFERENCES:**


1. JafarAbo Nada; Mohammad Rasmi Al-Mosa, 2018 International Arab Conference on Information Technology (ACIT), A Proposed Wireless Intrusion Detection Prevention and Attack System
2. Kinam Park; Youngrok Song; Yun-GyungCheong, 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigData Service), Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm
3. S. Bernard, L. Heutte and S. Adam "On the Selection of Decision Trees in Random Forests" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009, 978-1- 4244-3553-1/09/$25.00 ©2009 IEEE
4. A. Tesfahun, D. Lalitha Bhaskari, "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction" 2013 International Conferenceon Cloud & Ubiquitous Computing & Emerging Technologies, 978-0-4799-2235-2/13 $26.00 © 2013 IEEE
5. Le, T.-T.-H., Kang, H., & Kim, H. (2019). The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection. 2019 International Conference on Platform Technology and Service (PlatCon).Doi:10.1109/platcon.2019.8668960
6. Anish Halimaa A, Dr K.Sundarakantham:Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) 978-1-5386- 9439-8/19/$31.00 ©2019 IEEE "MACHINE LEARNING BASED INTRUSION DETECTION SYSTEM."
7. Mengmeng Ge, Xiping Fu, Naeem Syed, Zubair Baig, Gideon Teo, Antonio Robles- Kelly (2019). Deep Learning-Based Intrusion Detect ion for IoT Networks, 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), pp. 256-265, Japan.
8. R. Patgiri, U. Varshney, T. Akutota, and R. Kunde, "An Investigation on IntrusionDetection System Using Machine Learning" 978-1-5386-9276-9/18/$31.00 c2018IEEE.
9. Rohit Kumar Singh Gautam, Er. AmitDoegar; 2018 8th International Conferenceon Cloud Computing, Data Science & Engineering (Confluence) "An Ensemble Approach for Intrusion Detect ion System Using Machine Learning Algorithms."
10. Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahma, 2019International Conference on Robot ics, Electrical and Signal Processing Techniques(ICREST)"Network Intrusion Detect ionusing Supervised Machine Learning Technique with Feature Selection."
11. L. Haripriya, M.A. Jabbar, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)" Role of Machine Learning in Intrusion Detection System: Review"
12. Nimmy Krishnan, A. Salim, 2018 International CET Conference on Control,Communication, and Computing (IC4) " Machine Learning-Based Intrusion Detect ion for Virtualized Infrastructures"
13. Mohammed Ishaque, Ladislav Hudec,2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) "Feature extract ion using Deep Learning for Intrusion DetectionSystem."
14. Aditya Phadke, Mohit Kulkarni, Pranav Bhawalkar, Rashmi Bhattad, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)"A Review of Machine Learning Methodologies for Network Intrusion Detection."
15. Iftikhar Ahmad , Mohammad Basheri, Muhammad Javed Iqbal, Aneel Rahim, IEEE Access ( Volume: 6 ) Page(s): 33789 – 33795 "Performance Comparison of Support Vector Machine, Random Forest,and Extreme Learning Machine for Intrusion Detection."
16. B. Riyaz, S. Ganapathy, 2018 International Conference on Recent Trends in Advanced Computing (ICRTAC)" An Intelligent Fuzzy Rule-based Feature Select ion for Effective Intrusion Detection."