ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: **Wireless Sensor Network Attack Prediction Using AI**

Wireless Sensor Network Attack Prediction Using AI

Mani Mehala M[1], Mary Nisha D[2], Evelyn Tabitha E[3]

[1] PG Student, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

[2] Assistant Professor, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

[3] Assistant Professor, Department of Computer Science and Engineering, PET Engineering College, Tirunelveli, India

ABSTRACT

Wireless sensor network has attracted significant attention in research and development due to its tremendous applications in medical, military and defence, medical, environmental, industrial, infrastructure protection, and commercial applications to enable to interact with each other controlled remotely. A Wireless Sensor Network (WSN) has wide applications such as environmental monitoring and tracking of the target nodes for communication. The sensor nodes are equipped with wireless interfaces used for communication between the nodes and another network. Wireless Sensor Network suffers from many constraints that make security a primary challenge. When the sensor node is deployed in a communication environment unattended, the nodes are vulnerable to various attacks. The analysis of dataset by supervised machine learning technique(SMLT) to capture several information's like, variable identification, univariate analysis, bivariate and multivariate analysis, missing value treatments etc. A comparative study between machine learning algorithms had been carried out in order to determine which algorithm is the most accurate in predicting the type WSN attacks. The results show that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy, precision.

Keywords: Patient monitoring, Li-Fi, Sensors, LED, IoT, Light communication.

1 Introduction

Machine Learning is a system of computer algorithms that can learn from example through self- improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights. The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input and uses an algorithm to formulate answers. A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

learning to improve the user experience with personalizing recommendation. Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on. Machine Learning vs Traditional Programming Traditional programming differs significantly from machine learning. In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain. Traditional programming differs significantly from machine learning. In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical 2 foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

Figure 1: *Traditional Programming*

Machine learning is supposed to overcome this issue. The machine learns how the input and output data are correlated and it writes a rule. The programmers do not need to write new rules each time there is new data. The algorithms adapt in response to new data and experiences to improve efficacy over time.

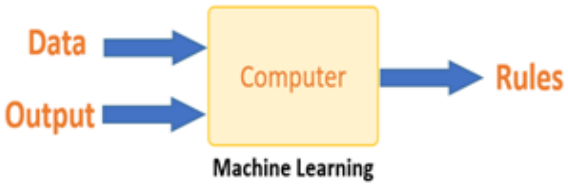


Figure 2: *Machine Learning*

How does Machine Learning Work? Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

unknown 3 situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict. The core objective of machine learning is the learning and inference. First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the data. One crucial part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a feature vector. You can think of a feature vector as a subset of data that is used to tackle a problem. The machine uses some fancy algorithms to simplify the reality and transform this discovery into a model. Therefore, the learning stage is used to describe the data and summarize it into a model.

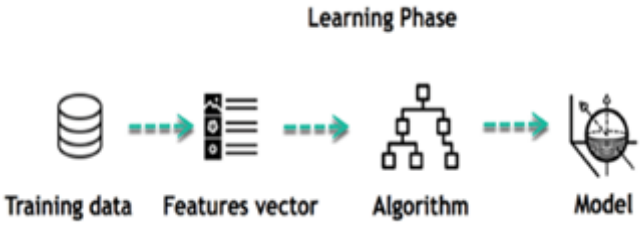


Figure 3: *Learning Phase*

For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant: This is the model Inferring When the model is built, it is possible to test how powerful it is on never-seen before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need 4 to update the rules or train again the model. You can use the model previously trained to make inference on new data.

The life of Machine Learning programs is straightforward and can be summarized in the following points: 1. Define a question 2. Collect data 3. Visualize data 4. Train algorithm 5. Test the Algorithm 6. Collect feedback 7. Refine the algorithm 8. Loop 4-7 until the results are satisfying 9. Use the model to make a prediction once the algorithm gets good at drawing the right conclusions, it applies that knowledge to new sets of data. 5.

Algorithms and where they are used? Machine learning Algorithms Machine learning can be grouped into two broad learning tasks: Supervised and Unsupervised. There are many other algorithms Supervised learning an algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans. You can use

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

supervised learning when the output data is known. The algorithm will predict new data. There are two categories of supervised learning: • Classification task • Regression task

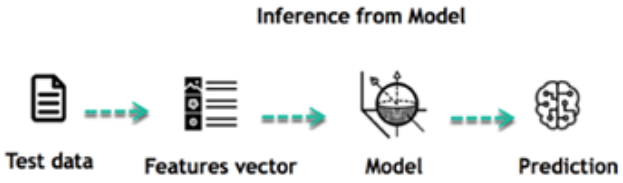


Figure 4: *Inference from Model*

6 1.2 Classification Imagine you want to predict the gender of a customer for a commercial. You will start gathering data on the height, weight, job, salary, purchasing basket, etc. from your customer database. You know the gender of each of your customer, it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features you have collected). When the model learned how to recognize male or female, you can use new data to make a prediction. For instance, you just got new information from an unknown customer, and you want to know if it is a male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this customer is a male, and 30% it is a female. The label can be of two or more classes. The above Machine learning example has only two classes, but if a classifier needs to predict object, it has dozens of classes (e.g., glass, table, shoes, etc. each object represents a class) Machine Learning (ML) algorithm: There are plenty of machine learning algorithms. The choice of the algorithm is based on the objective. In the Machine learning example below, the task is to predict the type of flower among the three varieties. The predictions are based on the length and the width of the petal. The picture depicts the results of ten different algorithms. The picture on the top left is the dataset. The data is classified into three categories: red, light blue and dark blue. There are some groupings. For instance, from the second image, everything in the upper left belongs to the red category, in the middle part, there is a mixture of uncertainty and light blue while the bottom corresponds to the dark category. The other images show different algorithms and how they try to classified the data. 7.

Challenges and Limitations of Machine Learning the primary challenge of machine learning is the lack of data or the diversity in the dataset. A machine cannot learn if there is no data available. Besides, a dataset with a lack of diversity gives the machine a hard time. A machine needs to have heterogeneity to learn meaningful insight. It is rare that an algorithm can extract information when there are no or few variations. It is recommended to have at least 20 observations per group to help the machine learn. This constraint leads to poor evaluation and prediction.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

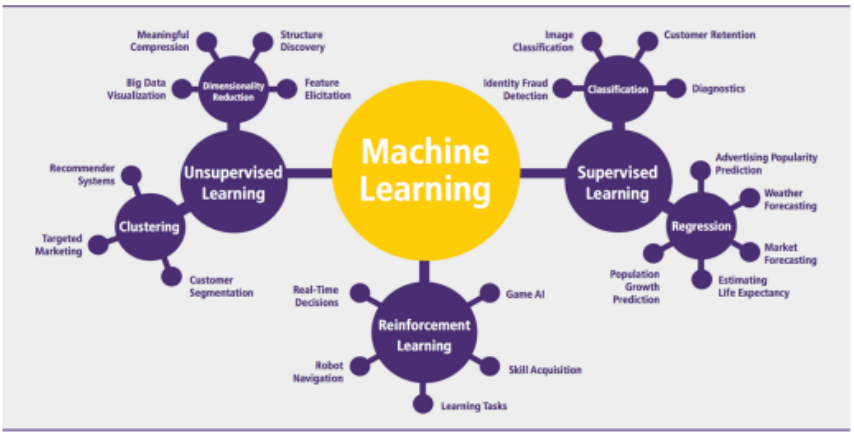


Figure 5: *Machine Learning Machine Learning*

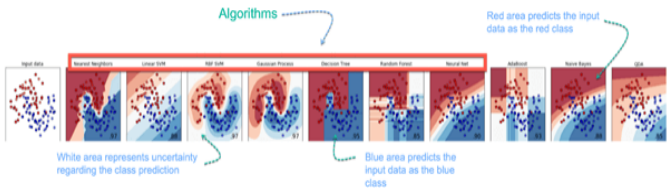


Fig.1.6 Algorithms

Figure 6: *Algorithms*

Application of Machine Learning Augmentation:

- Machine learning, which assists humans with their day-to-day tasks, personally or commercially without having complete control of the output. Such machine learning is used in different ways such as Virtual Assistant, Data analysis, software solutions. The primary user is to reduce errors due to human bias. Automation:
- Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots performing the essential process steps in manufacturing plants. Finance Industry
- Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud. 8 Government organization
- The government makes use of ML to manage public safety and utilities. Take the example of China with the massive face recognition. The government uses Artificial intelligence to prevent jaywalker. Healthcare industry

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

- Healthcare was one of the first industry to use machine learning with image detection. Marketing
- Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers develop advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, marketing department relies on AI to optimize the customer relationship and marketing campaign. Example of application of Machine Learning in Supply Chain Machine learning gives terrific results for visual pattern recognition, opening up many potential applications in physical inspection and maintenance across the entire supply chain network. Unsupervised learning can quickly search for comparable patterns in the diverse dataset. In turn, the machine can perform quality inspection throughout the logistics hub, shipment with damage and wear. For instance, IBM's Watson platform can determine shipping container damage. Watson combines visual and systems-based data to track, report and make recommendations in real-time. In past year stock manager relies extensively on the primary method to evaluate and forecast the inventory. When combining big data and machine learning, better forecasting techniques have been implemented (an improvement of 20 to 30 % over traditional forecasting tools). In term of sales, it means an increase of 2 to 3 % due to the potential reduction in inventory costs. 9 Example of Machine Learning Google Car For example, everybody knows the Google car. The car is full of lasers on the roof which are telling it where it is regarding the surrounding area. It has radar in the front, which is informing the car of the speed and motion of all the cars around it. It uses all of that data to figure out not only how to drive the car but also to figure out and predict what potential drivers around the car are going to do. What's impressive is that the car is processing almost a gigabyte a second of data. 1 Why is Machine Learning Important? Machine learning is the best tool so far to analyze, understand and identify a pattern in the data. One of the main ideas behind machine learning is that the computer can be trained to automate tasks that would be exhaustive or impossible for a human being. The clear breach from the traditional analysis is that machine learning can take decisions with minimal human intervention. Take the following example for this ML tutorial; a retail agent can estimate the price of a house based on his own experience and his knowledge of the market. A machine can be trained to translate the knowledge of an expert into features. The features are all the characteristics of a house, neighborhood, economic environment, etc. that make the price difference. For the expert, it took him probably some years to master the art of estimate the price of a house. His expertise is getting better and better after each sale.For the machine, it takes millions of data, (i.e., example) to master this art. At the very beginning of its learning, the machine makes a mistake, somehow like the junior salesman. Once the machine sees all the example, it got enough knowledge to make its

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

estimation. At the same time, with incredible accuracy. The machine is also able to adjust its mistake accordingly. Most of the big company have understood the value of machine learning and holding data. McKinsey have estimated that the value of analytics ranges from $9.5 trillion to $15.4 trillion while $5 to 7 trillion can be attributed to the most advanced AI techniques. Machine learning (ML) is the study of computer algorithms that improve automatically through experience.It is seen as a part of artificial intelligence. Machine 10 learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

2 Literature Survey

1) A Proposed Wireless Intrusion Detect ion Prevention and Attack System AUTHORS: JafarAbo Nada; Mohammad RasmiAl-Mosa This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet With the rapid deployment of wireless networks, the concept of network security has faced a lot of risks so it must provide security solutions. The classical methods of protecting networks from attacks are no longer adequate. For example, the intrusion detection system that works with wired networks has become useless with wireless networks. The Wireless technologies have opened a new field for network users. Because of its ease of use and setup, this technology has become popular and changing rapidly. However, the fear of the wireless world and the first threat is security. This is due to the nature of this network. With this increasing concern, it is necessary to start thinking about a security solution. This paper intends to propose a new wireless intrusion detection prevention and attack system to enhance the network security. Therefore, the paper will discuss the development of an intrusion detection system on wireless networks which is Wireless Intrusion Detection Prevention and Attack System “WIDPAS”. It is based on three main tasks: monitoring, analysis and defense. Through which it monitors denial of service attacks or false networks and then analyzes the attack and identifies the attacker and then protects the network users.

2) Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm AUTHORS: Kinam Park; Youngrok Song; Yun-Gyung Cheong In this paper, we present the results of our experiments to evaluate the performance of detecting different types of attacks (e.g., IDS, Malware, and Shellcode). We analyze the recognition performance by applying the Random Forest algorithm to the 12 various datasets that are constructed from the Kyoto 2006+ dataset, which is the latest network packet data collected for developing Intrusion Detection Systems. We conclude with discussions and future research projects.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

3) On the Selection of Decision Trees in Random Forests AUTHORS: S. Bernard, L. Heutte and S. Adam In this paper we present a study on the random forest (RF) family of ensemble methods. In aldquoclassicalrdquo RF induction process a fixed number of randomized decision trees are inducted to form an ensemble. This kind of algorithm presents two main drawbacks:

- The number of trees has to be fixed a priori (ii) the interpretability and analysis capacities offered by decision tree classifiers are lost due to the randomization principle. This kind of process in which trees are independently added to the ensemble, offers no guarantee that all those trees will cooperate effectively in the same committee. This statement rises two questions: are there any decision trees in a RF that provide the deterioration of ensemble performance? If so, is it possible to form a more accurate committee via removal of decision trees with poor performance? The answer to these questions is tackled as a classifier selection problem. We thus show that better subsets of decision trees can be obtained even using a sub-optimal classifier selection method. This proves that ldquoclassicalrdquo RF induction process, for which randomized trees are arbitrary added to the ensemble, is not the best approach to produce accurate RF classifiers. We also show the interest in designing RF by adding trees in a more dependent way than it is traditionally done in ldquoclassicalrdquo RF induction algorithms.

4) Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction AUTHORS: A. Tesfahun, D. LalithaBhaskari Intrusion Detection Systems (IDS) have become crucial components in computer and network security. NSL-KDD intrusion detection dataset which is an enhanced version of KDDCUP'99 dataset was used as the experiment dataset in this paper. Because 13 of inherent characteristics of intrusion detection, still there is huge imbalance between the classes in the NSL-KDD dataset, which makes harder to apply machine learning effectively in the area of intrusion detection. In dealing with class imbalance in this paper Synthetic Minority Over sampling Technique (SMOTE) is applied to the training dataset. A feature selection method based on Information Gain is presented and used to construct a reduced feature subset of NSL-KDD dataset. Random Forests are used as a classifier for the proposed intrusion detection framework. Empirical results show that Random Forests classifier with SMOTE and information gain based feature selection gives better performance in designing IDS that is efficient and effective for network intrusion detection.

5) The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection AUTHORS: Le, T.-T.-H., Kang, H., & Kim, H. A device or software appliance monitors a network or systems for malicious activity is an Intrusion Detection System (IDS). Conventional IDS does not detect elaborate cyber-attacks such as a low- rate DoS attack as well as unknown attacks. Machine Learning has attracted more and more interests in recent years to overcome these limitations. In this paper, we propose a novel method to improve intrusion detection accuracy of Gated Recurrent Unit (GRU) by embedding the proposed PCA-Scale with two options

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

including PCA-Standardized and PCA-MinMax into the layer of GRU. Both optional methods explicitly enforce the learned object feature maps by affecting the direction of maximum variance with positive covariance. This approach can be applied to GRU model with negligible additional computation cost. We present experimental results on two real-world datasets such as KDD Cup 99 and NSL-KDD demonstrate that GRU model trained with PCA-Scaled method achieves remarkable performance improvements.

3 Proposed System

The proposed model is to build a machine learning model for predicting wsn intrusion attacks. Previously they finds the accurate intrusion detection and isolation results only. wsn attack detection is an important technique for recognizing fraud activities, suspicious activities, network intrusion, and other abnormal events that may have great significance but are difficult to detect. The machine learning model is built by applying proper data science techniques like variable identification which is the dependent and independent variables. Each and every column‟s features are analyzed. Then the pre-processing and visualization of the data are done.The model is built based on the previous dataset where the algorithm learns data and gets trained different algorithms are used for better comparisons. The performance metrics are calculated and compared.

Advantages:

1) We are implementing the machine learning algorithm for classification purpose.
2) More than two machine learning algorithms are used for comparison of getting the best accuracy.
3) Deployment is done for getting result.

3.1 System Design

3.1.1 System Architecture

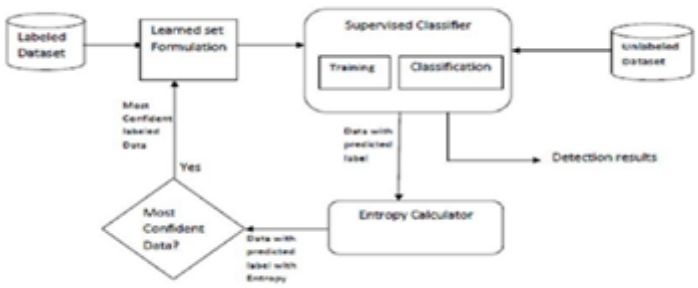


Figure 7: *System Architecture*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

3.2 Data Flow Diagram:

1) The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2) The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3) DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts 16 Input data Yes No information flow and the transformations that are applied as data moves from input to output. 4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

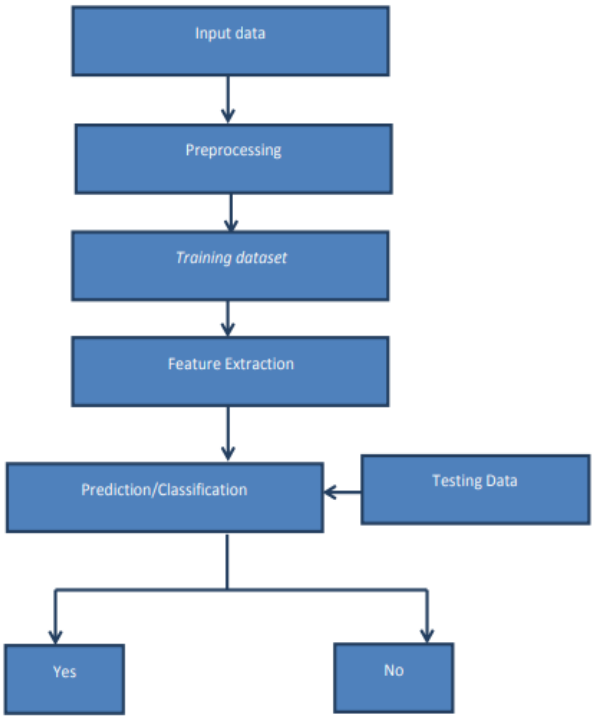


Figure 8: *Flow Diagram*

Preprocessing Feature Extraction Training dataset Prediction/Classification Testing Data 17

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

3.3 UML Diagrams

UML stands for Unified Modeling Language. UML is a standardized general purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

Goals

1) The Primary goals in the design of the UML are as follows:
2) Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
3) Provide extendibility and specialization mechanisms to extend the core concepts.
4) Be independent of particular programming languages and development process.
5) Provide a formal basis for understanding the modeling language.
6) Encourage the growth of OO tools market.
7) Support higher level development concepts such as collaborations, frameworks, patterns and components.
8) Integrate best practices 18 Preprocessing User Training Classification

3.2.1 Use Case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

ISSN (Online): 2583-7052

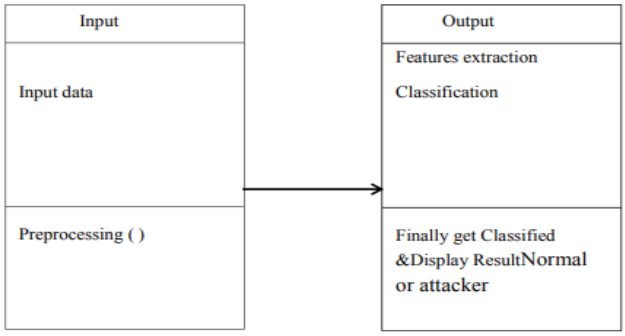
International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

Figure 9: *Input Data*

3.2.2 Class Diagram

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



3.2.3 Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams

3.2.4 Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

Modules

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

Modules Description Data Collection

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions and etc. The dataset used in this dataset taken from kdd Link: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html 5.3 Dataset: The dataset consists of 125974 individual data. There are 42 columns in the dataset, which are described below

Data Preparation: We will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.

Next we drop or remove all columns except for the columns that we want to retain. Finally we drop or remove the rows that have missing values from the data set. Split into training and evaluation sets Model Selection: The principal component analysis is the technique that is used, especially for the reduction of the dimension of the given dataset. The principal component analysis is one of the most efficient and an accurate method for reducing the dimensions of data, and it provides the desired results . This method reduces the aspects of the given dataset into a desired number of attributes called principal components. This method takes all the input as the dataset, which is having a high number of attributes so as the dimension of the dataset is very high. This method reduces the size of the dataset by taking the data points on the same axis. The data points are shifted on a single axis, and the principal components are carried out.

The PCA can be performed using the following steps:

- Take the dataset with all dimensions d.
- Calculate the mean vector for each dimension d.
- Calculate the covariance matrix for the wholedataset.
- Calculate the eigen vectors (e1, e2, e 3 … .ed), and eigen values (v1,, v2, v3,….vd).

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

- Perform sorting of eigenvalue in decreasing order and select n eigenvector with the highest eigenvalues to get a matrix of d*n= M.
- By using this M form a new sample space.
- The obtained sample spaces are the principal components. Random Forest is one of the most powerful methods that is used in machine learning for classification problems. The random forest comes in the category of the supervised classification algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the classifier.

Feature name	Description	Type
Duration	length (number of seconds) of the connection	continuous
Protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
Service	network service on the destination, e.g., http, telnet, etc.	discrete
Src_bytes	number of data bytes from source to destination	continuous
Dst_bytes	number of data bytes from destination to source	continuous
Flag	normal or error status of the connection	discrete
Land	1 if connection is from/to the same host/port; 0 otherwise	discrete
Wrong_fragment	number of "wrong" fragments	continuous
Urgent	number of urgent packets	continuous
Hot	number of "hot" indicators	continuous
Num_failed_logins	number of failed login attempts	continuous

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

Num_file_creations	number of file creation operations	continuous
Num_shells	number of shell prompts	continuous
Num_access_files	number of operations on access control files	continuous
Num_outbound_cmds	number of outbound commands in an ftp session	continuous
Is_hot_login	1 if the login belongs to the "hot" list; 0 otherwise	discrete
Is_guest_login	1 if the login is a "guest"login; 0 otherwise	discrete
Serror_rate	% of connections that have "SYN" errors	continuous
Serror_rate	% of connections that have "REJ" errors	continuous
Same_srv_rate	% of connections to the same service	continuous
Diff_srv_rate	% of connections to different services	continuous

1.Duration	length (number of seconds) of the connection
2.Protocol_type	type of the protocol, e.g. tcp, udp, etc.
3.Src_bytes	number of data bytes from source to destination
4.Dst_bytes	number of data bytes from destination to source
5.Is_hot_login	1 if the login belongs to the "hot" list; 0 otherwise
6.Is_guest_login	1 if the login is a "guest"login; 0 otherwise
7.Diff_srv_rate	% of connections to different services
8.Srv_diff_host_rate	% of connections to different hosts
9.Flag	normal or error status of the connection
10.Labels	Normal or attacker

Accuracy on test set

We got an accuracy of 99.1% on test set. Saving the Trained Model: Once you‟re confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle . Make sure you have pickle installed in your environment. Next, let‟s import the module and dump the model into .pkl file

3.3 System Testing

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

3.3.1 Types of Tests

Unit testing Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs.

All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive.

Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items: Valid Input : identified classes of valid input must be accepted. Invalid Input : identified classes of invalid input must be rejected. Functions : identified functions must be exercised. Output : identified classes of application outputs must be exercised. Systems/Procedures: interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined. System Test System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points. White Box Testing White Box Testing is a testing

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 1-18

Article Title: Wireless Sensor Network Attack Prediction Using AI

in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works. 6.1 Unit Testing: Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases. Test strategy and approach Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed. Features to be tested
- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page

Integration Testing Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error. Test Results: All the test cases mentioned above passed successfully. No defects encountered. 6.3 Acceptance Testing User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements. Test Results: All the test cases mentioned above passed successfully. No defects encountered.

4 Conclusion

As the involvement of the systems over the internet increasing rapidly, the security concerns have also seen. The proposed approach deals with the detection of intruders over the internet efficiently. The proposed algorithm has performed well as compared to the previously applied algorithms such as SVM, Naïve Bayes, and Decision Tree. The detection rates and the false error rates can be improved at a great extent by the proposed approach. The dataset used here

is the knowledge discovery dataset. The results obtained by our proposed method having the values for Performance time (min)is 3.24 minutes, Accuracy rate (%) is 96.78 %, and the Error rate (%) is 0.21 %.

Reference

1. JafarAbo Nada; Mohammad RasmiAl-Mosa, Year: 2018, "A proposed wireless intrusion detection prevention and attack system", 2018 International Arab Conference on Information Technology (ACIT).

2. Kinam Park; Youngrok Song; Yun-Gyung Cheong, Year: 2018, "Classification of attack types for intrusion detection systems using a machine learning algorithm", 2018 IEEE fourth international conference on big data computing service and applications (BigDataService).

3. Simon Bernard; Laurent Heutte; Sebastien Adam, Year: 2009, "On the selection of decision trees in random forests", In 2009 International joint conference on neural networks.

4. Abebe Tesfahun; D. Lalitha Bhaskari, Year: 2013, "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction", 2013 International Conference on Cloud &Ubiquitous Computing & Emerging Technologies.

5. Le, T.-T.-H., Hyeoun Kang; Howon Kim, Year: 2019, "The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection", 2019 International Conference on Platform Technology and Service (PlatCon).

6. AnishHalimaa A; Dr. K. Sundarakantham, Year: 2019, "Proceedings of the Third International Conference on Trends", in Electronics and Informatics (ICOEI 2019).