PREDICTION OF WSN-D ATTACK USING DATA SCIENCE TECHNIQUE

MR.M.GOUDHAMAN

PREETHI KANNAN

RANJITHA.C

Department Of Computer Science and Engineering

Jeppiaar Engineering College, Chennai, India

Abstract - *The main approach for this project is to predict the attack on wireless sensor network (WSN). This analysis aims to observe which features are most helpful in predicting the attack in wireless sensor network of blackhole, flooding, grayhole, normal and scheduling to see the general trends that may help us in model selection and hyperparameter selection. In this project, we have used machine learning classification methods to fit a function that can predict the discrete class of new input.*

Key words:

WSN, supervision algorithm, machine learning, logistic regression, ensemble algorithm, flask, data pre-processing, deployment.

1.Introduction:

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.

Wireless sensor networks refer to networks of spatially dispersed and dedicated sensors that monitor and record the physical conditions of the environment and forward the collected data to a central location. WSN faces high security threats considering most of them are deployed in unattended nature and hostile environment. In the aim of providing secure data processing in the WSN, many techniques are proposed to protect the data privacy while being transferred from the sensors to the base station. This work is focusing on attack detection which is an essential task to secure the network and the data.

An intrusion detection system is a device or software application that monitors a network or systems for malicious activity or policy violations. An IDS is one possible *solution to address a wide range of security attacks in WSNs*

2.Literarure survey:

A. Title: Evaluation of Machine Learning Algorithms for Intrusion Detection System in WSN

This paper addresses the security issues in WSN by establishing potential automated solutions for identifying associated risks. It also evaluates the effectiveness of various machine learning algorithms on two types of datasets, mainly, KDD99 and WSN datasets. The aim is to analyse and protect WSN networks in combination with Firewalls, Deep Packet Inspection (DPI), and Intrusion Prevention Systems (IPS) all specialized for the overall protection of WSN networks.

3.Scope Of the project:

The scope of this project is to investigate a dataset of network connection attacks for Routing records for using machine learning technique. To identifying network connection is attacked or not. This analysis aims to observe which features are most helpful in predicting the WSN of blackhole, flooding, grayhole, normal and scheduling to see the general trends that may help us in model selection and hyper parameter selection. To achieve used machine learning classification methods to fit a function that can predict the discrete class of new input. The projects goal is gather the different types to data, pre-processing the data, train the data's to a particular modal, test the modal and predict the type of attack. Intrusion detection system can predict the type of attack occurred in the wireless sensor network.

4.System requirement:

Software Requirements:

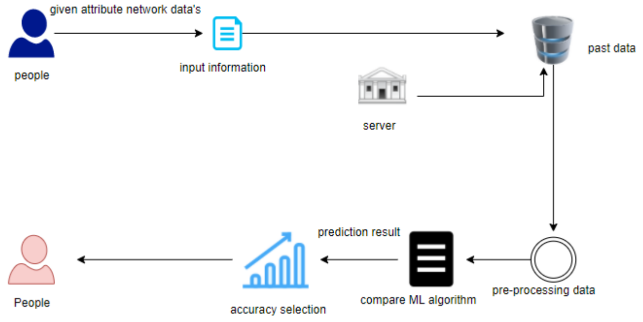
- *Operating System: Windows*
- *Tool: Anaconda with Jupyter Notebook*

Hardware requirements:

- *Processor: Pentium IV/III*
- *Hard disk: minimum 80 GB*
- *RAM: minimum 2 GB*

5.Design architecture

System architecture



6.Proposed system:

In proposed system, there are 4 modules. They are:

- *Data pre-processing (Module-01)*
- *Data Analysis of Visualization (Module-02)*
- *Comparing Algorithm with prediction in the form of best accuracy result (Module-03)*
- *Deployment Using Result (Module-04)*
- ***Module 1: Data pre-processing***

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. To achieving better results from the applied model in Machine Learning method of the data has to be in a proper manner. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making

Module II: Data Analysis of Visualization

Data visualization is an important skill in applied statistics and machine learning. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts. In data pre-processing, we will be finding the missing value, duplicate value and description of data type whether it is float variable or integer. Data cleaning / preparing by rename the given dataset and drop the column etc. The aim is to analyze the uni-variate, bi-variate and multivariate process.

Module III: Comparing Algorithm

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. Each model will have different performance characteristics. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. A way to do this is to use different visualization methods to *show the average accuracy, variance and other properties of the distribution of model accuracies.4 different algorithms are used for comparison:*

- *Logistic regression*
- *Random forest classifier*
- *Navie Bayes*
- *Decision tree classifier*

Module III: Deployment

Flask (Web Frame Work)

Flask is an API of Python that allows us to build up web-applications. It was developed by Armin Ronacher. Flask framework is more explicit than Django framework and is also easier to learn because it has less base code to implement a simple web-Application. In this project, flask framework calls all the AI files from the different module. This will help

us to compare the algorithm and predicts the output which has more accuracy. Here all the AI files are being called from the different module.

6.Advantages:

1.The main advantage in this project is the usage of ensemble algorithm

2. We calculate the Confusion Matrix

3. We are Deploying in Flask Framework

4. The anomaly detection can be automated process using the machine learning.

5. Performance metric are compared in order to get better model.

7.Conclusion:

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score will be find out by comparing each algorithm with type of all WSN Attacks for future prediction results by finding best connections. This brings some of the following insights about diagnose the network attack of each new connection. To presented a prediction model with the aid of artificial intelligence to improve over human accuracy and provide with the scope of early detection. It can be inferred from this model that, area analysis and use of machine learning technique is useful in developing prediction models that can helps to network sectors reduce the long process of diagnosis and eradicate any human error.

RESULT:

The result of this project is to investigate a dataset of network connection attacks for Routing records for medical sector using machine learning technique. To identifying network connection is attacked or not. Exploration data analysis of variable identification of loading the given dataset, Import required libraries packages, Analyze the general properties, find duplicate and missing values, Checking unique and count values, uni-variate data analysis**,** rename, add data and drop the data, to specify data type.

8.FUTURE WORK:

Network sector want to automate and detecting the attacks of packet transfers from eligibility process (real time) based on the connection detail. To automate this process by show the prediction result in web application or desktop application at cloud. To optimize the work to implement in Artificial Intelligence environment.

REFERENCE:

[1] I. F. Akyildiz and M. C. Vuran, "Wireless sensor networks," John Wiley & Sons, vol. 4, 2010.

[2] A. H. Farooqi and F. A. Khan, “Intrusion Detection Systems for Wireless Sensor Networks: A Survey, ”In International Conference on Future Generation Communication and Networking, Springer, Berlin, Heidelberg, pp. 234–241, 2009.

[3] Ghosal, A., & Halder, S., "Intrusion detection in wireless sensor networks: Issues, challenges and approaches, "In Wireless networks and security. Springer, pp. 329