Intrusion Detection for HealthCare Network using Machine Learning

Sahana B.G[1], Shriya Urankar[2], Rashmi T.V[3]

[1]*Student, Dept. Of Information Science and Engineering, BNMIT, Karnataka, INDIA*
[2]*Student, Dept. Of Information Science and Engineering, BNMIT, Karnataka, INDIA*
[3]Asst.Professor, *Dept. Of Information Science and Engineering, BNMIT, Karnataka, INDIA*

---***---

Abstract - *An intrusion is any activity that is designed to compromise data security. This can be through more menacing and pervasive formats like ransomware or unintentional data breaches by employees or others connected to the network. An intrusion may include DDOS attacks, cyber-enabled equipment destruction, accidental employee security breaches, untrustworthy users and social engineering attacks. Health information is the brief and precise history of a patient's life and ailment history. From the medical perspective, it is a collection of recorded information about a particular patient. Proper management of patient health information defines the quality of healthcare, therefore a streamlined health care system completely depends upon a good health information storage and preservation system. With the increase in threats, there have been several attempts to build an effective intrusion detection system and the aim is to build a system which can efficiently detect intrusions and provide safety.*

***Key Words*: Intrusion Detection, HealthCare, Patient, Security, Attacks, Machine Learning.**

1. INTRODUCTION

Over the last few years, machine learning techniques for intrusion detection have become prevalent in order to minimize and control security breaches and prevent attacks. Modern technologies such as robotics, computer vision and Artificial Intelligence (AI) are being employed currently in the healthcare environment. In the recent years, healthcare industries are depending on machine learning techniques to tackling security challenges in healthcare. However, there are various methods for Intrusion Detection System (IDS) which utilizes machine learning. Machine learning technique is put into service because of its continually evolving diversity and for its ability to accomplish a high rate of false-positive traffic with a less computational cost. Many IDS classification techniques were proposed by several researchers to sort out network traffic into benign and malignant. Machine learning algorithms play a vital role in the cybersecurity domain. Machine learning algorithms such as Convolutional Neural Networks (CNN), Decision Tree (DT), K-nearest neighbor (K-NN), support vector machine (SVM) have been integrated with intrusion detection systems which aid in improving the classification results. Though Healthcare networks provide a wide range of opportunities they also have a set of obstacles including privacy concerns and serious security issues since healthcare systems contain sensitive and life-critical medical data. Research in this field shows that the healthcare industry is more vulnerable to cybersecurity attacks in the recent time. The Intrusion detection system (IDS) is more efficient and convenient for hospitals. The access of data is made more convenient for the authorized users. The admin gets an alert from the IDS about unauthorized users. The detection system is automated and hence, human interference is reduced. The machine is well trained with suitable algorithms, making the system more productive.

2. PROBLEM STATEMENT

The increase in the attacks on healthcare networks has been a cause of concern, since the sensitive data of the patients can be acquired by malicious users. Network security attacks have been a challenge that many are trying to solve.

3. METHODOLOGY

The project has been implemented by dividing the entire project into three modules. They are:

- Pre-processing
- Training
- Prediction

3.1 Pre-processing

The dataset is loaded and printed along with the dataset field names after setting up the environment. The field's names include duration, protocol type, service, logged_in and so on. Mapping of the attack fields to attack classes such as DoS, Probe, U2R, R2L is done. As a result, the field name attack is replaced with attack_class. The Dataset is described, wherein the mean, standard deviation, minimum and so on is found out. This helps in further analysis of the dataset along with identifying the features which are of importance.

The frequency and the percentage of attack class distribution in the dataset is found. A graph is plotted accordingly. The attributes containing integer and float data types are taken. These are then set to a specific range. Based on scaling, each attribute is assigned a numerical value, making it easier to refer and access the attribute. Similarly, scaling is done for string data types, which are assigned numerical values.

3.2 Training the Model

The website is developed exclusively for the hospital network. This includes the python model which is trained using the NSL-KDD dataset under various conditions. The total number of 1,48,000 inputs are considered from the dataset. The model is made to learn to classify the inputs into normal or malicious.

Pickle API is used for loading the model as a file and is also used as a connection mechanism between the client and server side. Couple of metrics needs to be created to determine how accurate the model is. Then, all the information will be compiled for the model so that it is ready to be trained.

3.3 Prediction

For prediction confusion matrix of the input is given to the model and prediction of the same is given. The input values which include logged_in and count are allowed to predict for the normal and malicious attack types and the trained Decision Tree Classifier (DTC) is applied which is double exponential and turns values into probabilities. The result is obtained in the form of categorical value and the output predicted will be normal or malicious.

Figure 1 shows the system architecture diagram with different components of the proposed system.

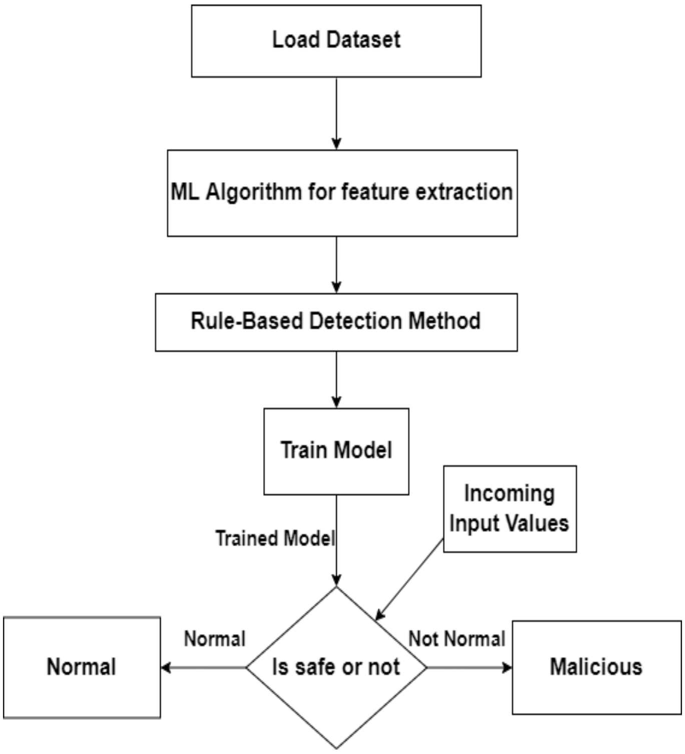


Figure 1: Architecture Diagram of the System.

3.4 Algorithm

Algorithm: Decision Tree

Description: Detects the type of attack, by considering login and count features.

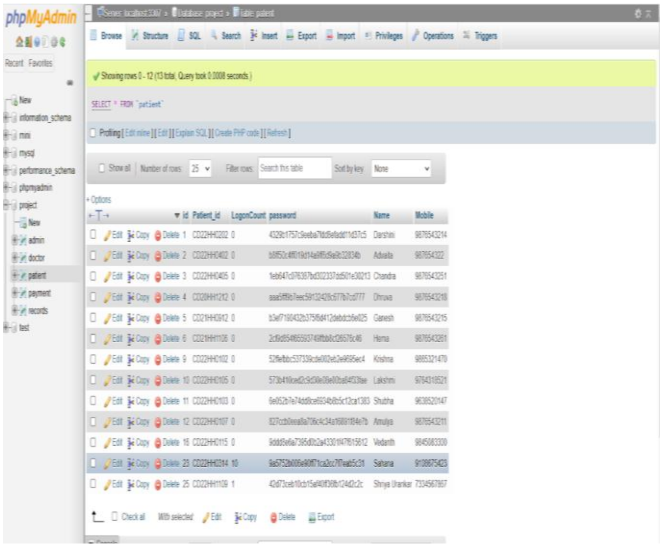
Input: input.csv

Output: Output gets saved in the output file, output.txt.

```
Step 1: Read the input from input.csv file.
Step 2: Feed the login and count values as input to the DTC model.
Step 3: Repeat for all the values detected:
        Step 3.1: Obtain the detected values
        Step 3.2: Feed the obtained values to the DTC model
        Step 3.3: Compare the value obtained with the dataset, then,
                Step 3.3.1: Classify the attack type
                Step 3.3.2: Feed the obtained result to the output file
Step 4: Return the result.
```

4. RESULTS

Figure 2 shows the count information of the patient after the patient has logged in. The count information specifies the number of connections to the same host as current connection in past two seconds.



Showing rows 0 - 12 (13 total, Query took 0.0008 seconds.)

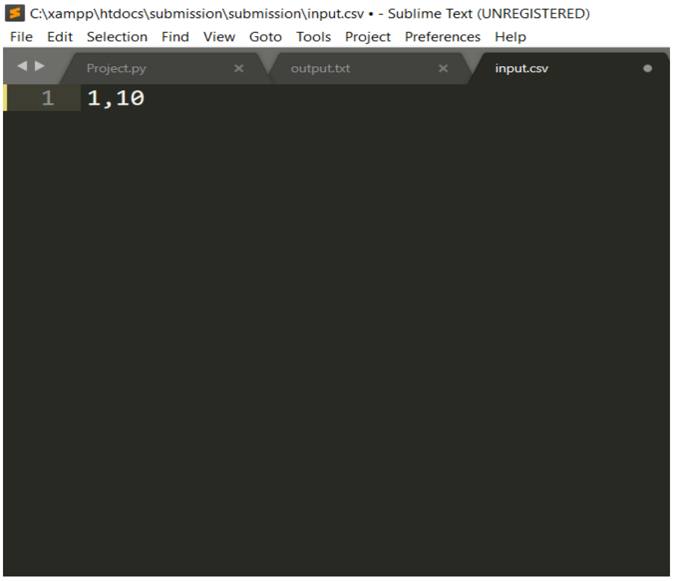
```
SELECT * FROM `patient`
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

id	Patient_id	LogonCount	password	Name	Mobile
1	CO22HH0202	0	[illegible]	Darshini	9876543214
2	CO22HH0402	0	[illegible]	Advaita	987654322
3	CO22HH0405	0	[illegible]	Chandra	9876543251
4	CO20HH1212	0	[illegible]	Dhruva	9876543218
5	CO21HH0912	0	[illegible]	Ganesh	9876543215
6	CO21HH1106	0	[illegible]	Hema	9876543281
9	CO22HH0102	0	[illegible]	Krishna	9865321470
10	CO22HH0105	0	[illegible]	Lakshmi	9764318521
11	CO22HH0103	0	[illegible]	Shubha	9638520147
12	CO22HH0107	0	[illegible]	Amulya	9876543211
16	CO22HH0115	0	[illegible]	Vedanth	9845083330
23	CO22HH0314	10	[illegible]	Sahana	9108675423
25	CO22HH1109	1	[illegible]	Shriya Urankar	7334567897

Figure 2: Patient Data Table

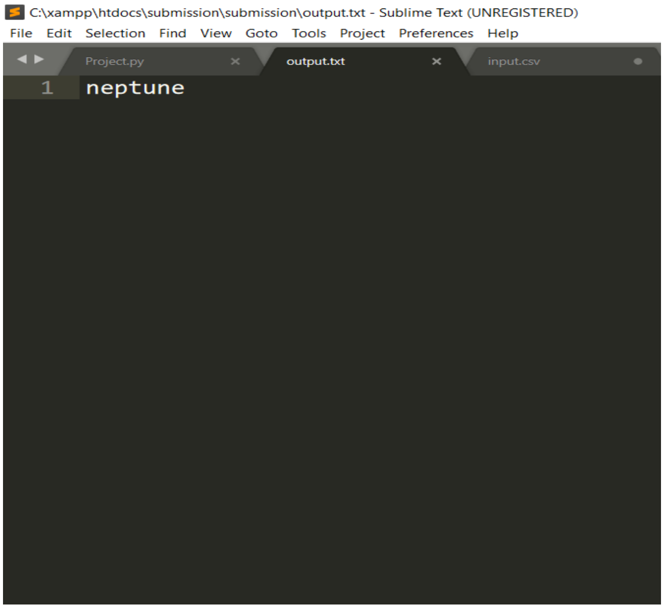
Figure 3 shows the input values for Neptune attack, which contain the logged_in and count information. Neptune attack is a type of Denial of Service (DoS) attack which is accomplished by sending multiple requests to the targeted machine in order to overload the systems and prevent some or all requests from being fulfilled.


C:\xampp\htdocs\submission\submission\input.csv • - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help
Project.py
output.txt
input.csv

```
1,10
```

Figure 3: Input for Neptune attack

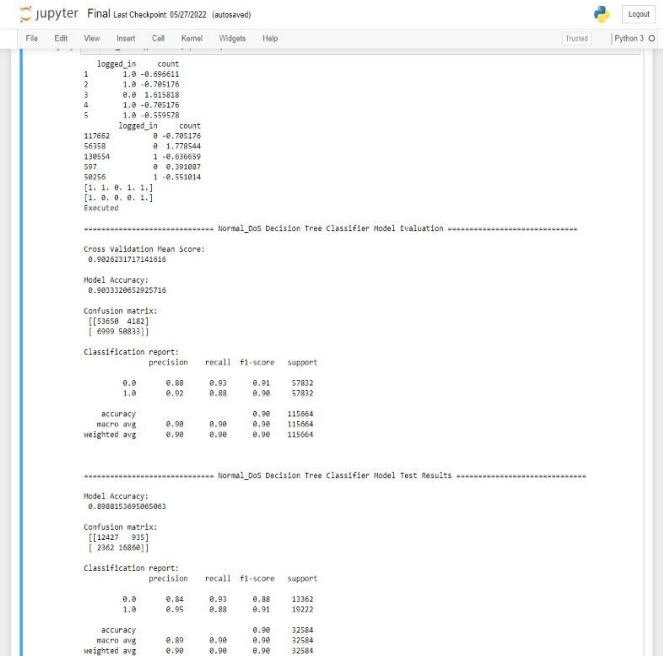
Figure 4 shows the output values for Neptune attack, which contain the output for the input given in figure 3.


C:\xampp\htdocs\submission\submission\output.txt - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help
Project.py
output.txt
input.csv

```
neptune
```

Figure 4: Output for Neptune attack

Figure 5 shows the calculated accuracy for the trained model. The accuracy calculated is 90.3%. The model accuracy is calculated using cross validation mean score and the confusion matrix along with classification report is displayed.


jupyter Final Last Checkpoint: 05/27/2022 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3

```
   logged_in     count
1        1.0 -0.696611
2        1.0 -0.705176
3        0.0  1.615818
4        1.0 -0.705176
5        1.0 -0.559578
        logged_in     count
117662          0 -0.705176
56358           0  1.778544
130554          1 -0.636659
597             0  0.391087
50256           1 -0.551014
[1. 1. 0. 1. 1.]
[1. 0. 0. 0. 1.]
Executed

============================== Normal_DoS Decision Tree Classifier Model Evaluation ==============================

Cross Validation Mean Score:
 0.9026231717141616

Model Accuracy:
 0.9033320652925716

Confusion matrix:
 [[53650  4182]
 [ 6999 50833]]

Classification report:
               precision    recall  f1-score   support

         0.0       0.88      0.93      0.91     57832
         1.0       0.92      0.88      0.90     57832

    accuracy                           0.90    115664
   macro avg       0.90      0.90      0.90    115664
weighted avg       0.90      0.90      0.90    115664


============================== Normal_DoS Decision Tree Classifier Model Test Results ==============================

Model Accuracy:
 0.8988153695065063

Confusion matrix:
 [[12427   935]
 [ 2362 16860]]

Classification report:
               precision    recall  f1-score   support

         0.0       0.84      0.93      0.88     13362
         1.0       0.95      0.88      0.91     19222

    accuracy                           0.90     32584
   macro avg       0.89      0.90      0.90     32584
weighted avg       0.90      0.90      0.90     32584
```

Figure 5: Accuracy Calculated

5. CONCLUSIONS

The project has been a great learning process in the field of Machine Learning and its applications. Intrusion Detection for HealthCare Network using Machine Learning is built to help healthcare industry to tackle intrusions.

Intrusion detection for a vast healthcare network is a difficult task, but a necessary step forward into machine intelligence in order to enable patients to rely on these networks. Classification of attack classes on a large dataset is challenging than classification on a smaller dataset, but is essential. The proposed system aims to provide users with a novel, reliable hospital network which can ensure the security of the patients and their data by efficiently classifying the attacks.

6. FUTURE ENHANCEMENT

The proposed work can be enhanced by including all the 42 features of the dataset as well as providing a mechanism to immediately alert the admin regarding the attack summary. Admin can also be given the authority to allow or block the users from accessing the network. The website can also be made multi-lingual.

REFERENCES

[1] Anar A Hady, Ali Ghubaish, Tara Salman, Devrim Unal, and Raj Jain, "Intrusion Detection System for Healthcare Systems Using Medical and Network Data", IEEE Access/ ACCESS.2020.3000421, vol. 8, June 18, 2020.

[2] Geethapriya Thamilarasu, Adedayo Odesile, And Andrew Hoang, "An Intrusion Detection System for Internet of Medical Things", IEEE Access/ACCESS.2020.3026260, vol. 8, October 14, 2020.

[3] Sudhakar Sengan, Osamah Ibrahim Khalaf, Vidya Sagar P, Dilip Kumar Sharma, Arokia Jesu Prabhu L, Abdulsattar Abdullah Hamad, "Secured and Privacy-Based IDS for Healthcare Systems on E-Medical Data Using Machine Learning Approach", International Journal of Reliable and Quality E-HealthcareVolume 11 • Issue 3, November 14, 2020.

[4] Seresane V, Krishna Reddy, Jayanthi K Murthy, "Intrusion Detection in Hospital Automation using Internet of Things (IoT)", Turkish Journal of Physiotherapy and Rehabilitation; 32(2)ISSN 2651-4451 | e-ISSN 2651-446X, December 20, 2020.

[5] Tich Phuoc Tran, Pohsiang Tsai, Tony Jan and Xiaoying Kong, "Network Intrusion Detection using Machine Learning", SBN: 978-953-307-033-9,InTech, December 27, 2020.

[6] Dr. G Umarani Srikanth, Priyadharsini S, "Prediction Of Network Attacks using Machine Learning Techniques", International Journal of Engineering Applied Sciences and Technology, 2021 Vol. 5, Issue 10, ISSN No. 2455-2143, Pages 112-118, February 20, 2021.

[7] Celestine Iwendi, Joseph Henry Anajemba, Cresantus Biamba and Desire Ngabo, "Security of Things Intrusion Detection System for Smart Healthcare", electronics10121375, June 8, 2021.

[8] Tanzila Saba - Prince Sultan University, Riyadh, Saudi Arabia, "Intrusion Detection in Smart City Hospitals using Ensemble Classifiers", IEEE Explore, July 6, 2021, 978-1-6654-2238-3/20 ©2020 IEEE.