International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 5, May 2020||

Intrusion Detection System using Machine Learning

Namish Khanna[1], Harnoor Singh[2], Aditya Sharma[3]

B.E. Student, Department of Computer Science and Engineering,Chandigarh University, Gharuan, Punjab, India[1]

B.E. Student, Department of Computer Science and Engineering,Chandigarh University, Gharuan, Punjab, India[2]

B.E. Student, Department of Computer Science and Engineering,Chandigarh University, Gharuan, Punjab, India[3]

ABSTRACT: The Purpose of making this project is to Introduce the User to Intrusion Detection System. i.e. How one can find whether the attack was happened on a particular website or not. Intrusion Detection System helps to find whether there is any malicious activity happening to website or not. Using the Random Forest Algorithm with Machine Learning, we tend to identify and find those websites which are being attacked. As we all are turning into digital world so, it was difficult for everyone to tell whether his/her website is safe or not. Hackers uses different methods to attack on our data, and our system helps to predict that whether there is anything wrong or not in more efficient manner.

KEYWORDS: Random Forest Algorithm, Machine Learning, Extraction.

I. INTRODUCTION

The purpose of this project is to build an Intrusion detection system using Machine learning to predict and detect the various kinds of attack. Though there are manual ways to detect this, but the model purposed by us can easily detect the attacks without much of human efforts. Moreover, the model predicts the values accurately up to 99% which is far better than a human being. This can be done by using the Random Forest algorithm.

With further tuning of the parameters of the algorithm we are able to achieve such great results. From six algorithms the results were compared and finally the Random Forest algorithm was chosen. The front-end of the project has been made using HTML, CSS and Bootstrap framework (as it is a web-based project), and the backend is made in Django framework.

An IDS will be a bit of put in software system or a physical appliance that monitors network traffic so as to notice unwanted activity and events like outlaw and malicious traffic, traffic also violates security policy, and traffic that violates acceptable use policies. We are going to predict the bot attacks which happened on different websites and using that dataset we are going to know the prediction with an accuracy of 99% using random forest algorithm for better results.

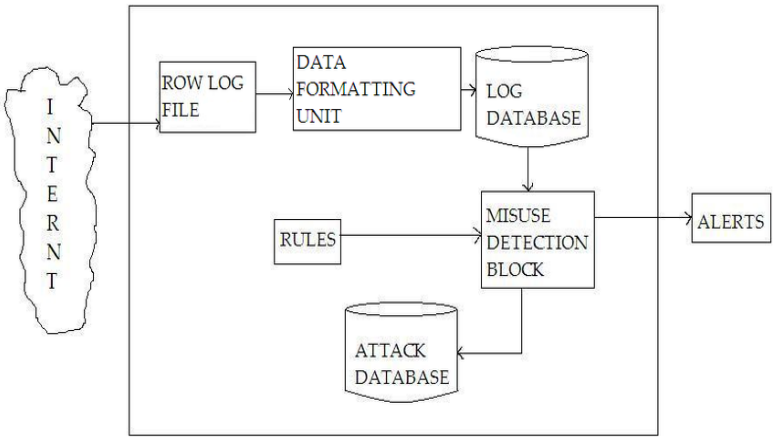


Figure 1: Flow diagram

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 5, May 2020||

II. RELATED WORK

According to work done by analysis the Deep packet review may be a major element in Network Intrusion Detection systems (NIDSes), the incoming information stream of the packets ought to be compared with patterns in associate attack info, byte-by-byte, mistreatment string matching or regular expression matching.

Regular expression matched, despite flexibility/efficiency within the attack identification, has the very best computation and storage complexities for NIDSes, creating line-rate packet process difficult. Stride Finite Automata (StriFA), a replacement finite automata family, to accelerate string matching and regular expression matching was given by Wang, et al., (2013).

Altering from the ancient finite automata, that scan a complete traffic stream to find malicious data, StriFA scans solely partial traffic stream to placed suspicious data. StriFA technique was enforced in code and evaluated on varied traces.Simulation showed that Stria acceleration offers magnified speed over ancient nondeterministic finite automaton and settled finite automaton and reduces memory demand at the same time.

III. PROPOSED WORK

In the previous research or review papers only, the prediction is done on very small datasets and even with different algorithms which lead to low accuracy but in this project, we have used Random Forest Algorithm which gives us accuracy of above 95% which was quite good.

We have used dataset of more than one lakh values for training as well as test so that we will be able to predict it in a more good manner. We have used many other algorithms in order to check whether we can get more accuracy but we were not able to get more accuracy than the random forest algorithm.

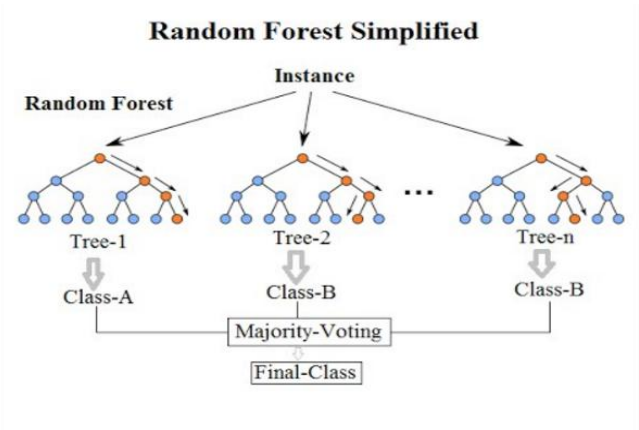


Figure 2. Random Forest Algorithm

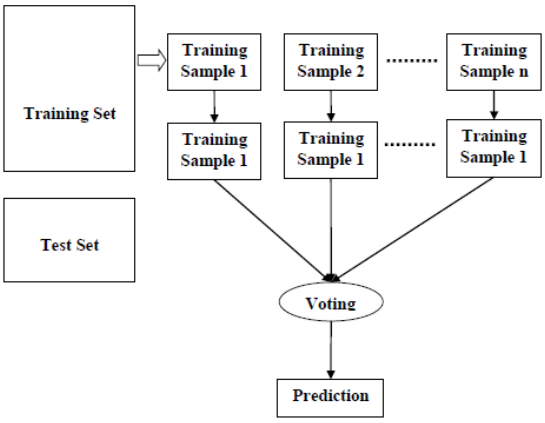


Figure 3. Training Module

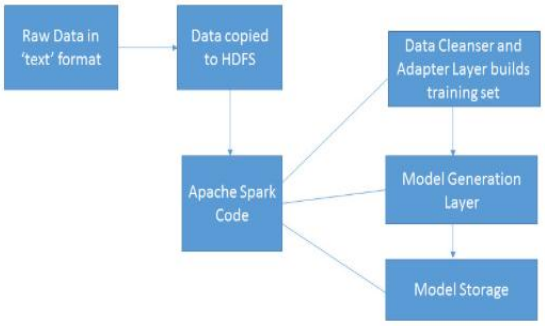


Figure 4. Flow Chart

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 5, May 2020||

IV. METHODOLOGY

a. Learn which Technology will be used by Machine Learning.
b. Download the Dataset and then Convert it into .CSV and .PKL format if they were in another format.
c. Remove the Unwanted Rows whose values are missing or NULL and then update the rows and columns accordingly.
d. HTML, CSS and JavaScript is included in order to print the result in more specified way.
e. In Machine Learning, there are various methods to deal with missing values and then optimize it.
f. Creation of Heatmaps helps to find which column to select for prediction so that we can get high accuracy.
g. Various algorithms like linear regression, Multiple Regression, Decision Tree, Random forest etc. were used in order to find the result.
h. Frontend and Backend was connected in Django in order to optimize the output.
i. After that we combine all the files accordingly and then find which algorithm to select for prediction.
j. At last we selected the Random Forest Algorithm as we are getting the highest accuracy with this algorithm.
k. After all this we were ready to predict the result and tell whether it was a bot attack or not.

A. Dataset

The dataset of more than one lakh values was taken in order to predict the results. The dataset is converted into pickle format so that it can be handled more effectively and efficiently.

B. Used Models

We have used random forest algorithm for the prediction of data in efficient manner and with high accuracy. Random Forest Algorithm is a Supervised Learning Algorithm which we tend to select data in large amount to increase the accuracy.

C. Experimental setting

Firstly, we convert the dataset in pickle format so that it would be easy to handle and then we took 60% of dataset for training purpose of our model and 40% for the testing purpose of our model for prediction.

Table 1: Different Algorithm Results

S. No.	Algorithm	Accuracy
1.	Linear Regression	65.76%
2.	Decision Tree	82.34%
3.	Random Forest	98.87%

We have tested on these Algorithms and got the respective results. Random Forest Algorithm was best from all other Algorithms and gave the highest accuracy.

V. EXPERIMENTAL RESULTS

A column is chosen from the dataset at the runtime i.e. dynamically and is shown on the webpage. The user can select any of the desired column and see what the result is. The model takes the column values of the data and then apply the 'Random Forest' and returns the result back. The predicted result is accurate up to accuracy of 99%.

The Objective to build this project is to find whether the attack was happened on the particular site or not. We have found different parameters through which we found that on which parameters the attack has happened and on which parameters attack has not happened. The main motive to build this website is just to help the users and tell them whether the attack was bot or benign.

User can now easily predict that whether the particular website is under bot attack or not.

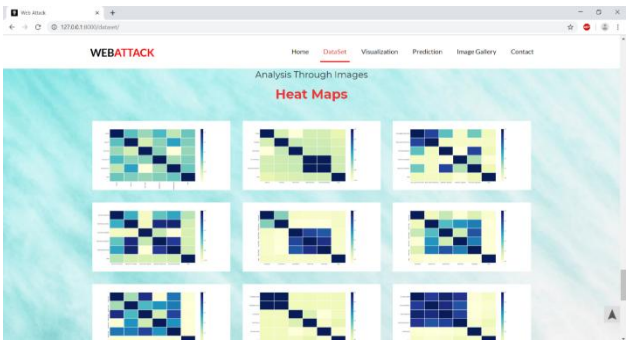


Figure 5. Heat Maps

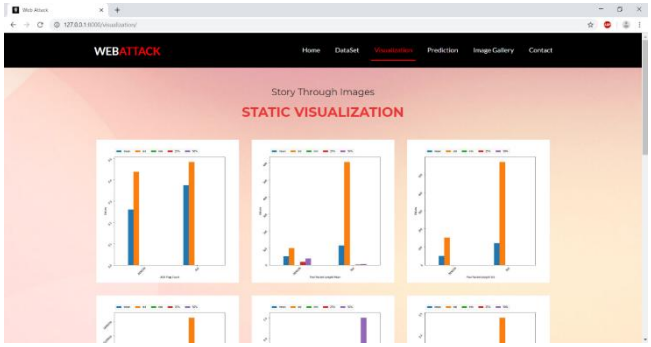


Figure 6. Static Visualization

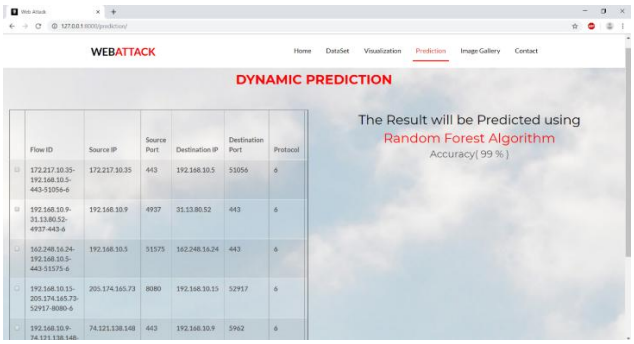


Figure 7. Dynamic Prediction Before

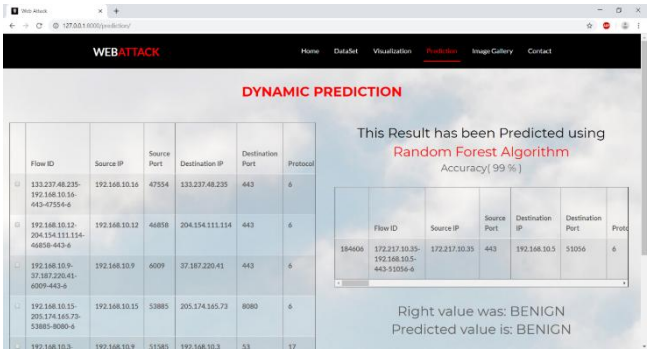


Figure 8. Dynamic Prediction After

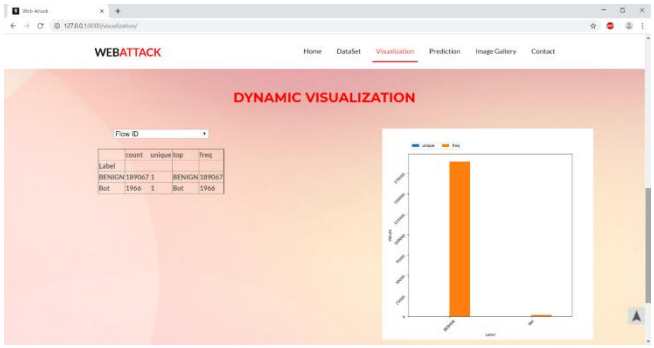


Figure 9. Dynamic Visualization

```
from django.shortcuts import render
from django.http import HttpResponse
import pandas as pd
import matplotlib.pyplot as plt
import io
from PIL import Image, ImageDraw
import PIL, PIL.Image
from io import BytesIO
import base64
import random
import pickle
from bs4 import BeautifulSoup as bsu
from sklearn.ensemble import RandomForestRegressor
import numpy as np
from django.core.mail import send_mail
from django.conf import settings

```

Figure 10. Used Libraries

The Figure 5. Shows Heats maps which describes us that which columns we have to take in order to predict our result in mor accurate way so by selecting right columns we will get more accuracy.

The Figure 7. And figure 8. Shows us the dynamic prediction in which we will select the value from list and then it will tend to predict the value that whether it was bot attck or not and then it will show us the result.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 5, May 2020||

VI. CONCLUSION

The Objective to build this project is to find whether the attack was happened on the particular site or not. We have found different parameters through which we found that on which parameters the attack has happened and on which parameters attack has not happened. Being safe now-a-days is very difficult but our system helps to predict the odds which it detects to safe your website from hackers.

With the help of Machine Learning we not only can find the attacked website but can also prevent them from happening in future by adding some preventions in the Website.

The future scope of this project is we are dealing it with only bot attack and in future we will add more attacks and try to predict those attack and help the users to check attack in more ways. Different attacks involvedness helps us to predict it in more efficient manner.

REFERENCES

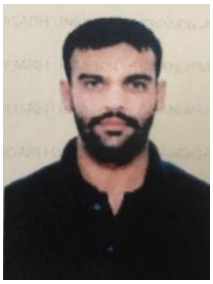
[1] Ashfaq, Rana Aamir Raza, et al. "Fuzziness based semi-supervised learning approach for intrusion detection system." *Information Sciences* 378 (2017): 484-497.
[2] Aljawarneh, Shadi, Monther Aldwairi, and Muneer Bani Yassein. "Intrusion detection system through feature selection analysis based anonmaloy and building hybrid efficient model." *Journal of Computational Science* 25 (2018): 152-160.
[3] Aburomman, Abdulla Amin, and Mamun Bin Ibne Reaz. "A novel SVM-kNN-PSO ensemble method for intrusion detection system." *Applied Soft Computing* 38 (2016): 360-372.
[4] Ambusaidi, Mohammed A., et al. "Building an intrusion detection system using a filter-based feature selection algorithm." *IEEE transactions on computers* 65.10 (2016): 2986-2998.
[5] Farnaaz, Nabila, and M. A. Jabbar. "Random forest modeling for network intrusion detection system." *Procedia Computer Science* 89.1 (2016): 213-217.
[6] https://ieeexplore.ieee.org/abstract/document/7124898
[7] https://ieeexplore.ieee.org/abstract/document/8399291
[8] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, "Internet of things: A survey on enabling technologies protocols and applications", *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2347-2376, 2015.
[9] S. Keele, "Guidelines for performing systematic literature reviews in software engineering", *Technical Report Ver. 2.3 EBSE Technical Report. EBSE*, 2007.
[10] S. Horrow, S. Anjali, "Identity Management Framework for Cloud Based Internet of Things", *SecurIT '12 Proceedings of the First International Conference on Security of Internet of Things*, pp. 200-203, 2012.
[11] U. Oktay, O. K. Sahingoz, "Proxy Network Intrusion Detection System for the Cloud Computing Software", pp. 98-104, 2013, ISBN 978-1-4673-5613-8.
[12] A. Patel, M. Taghavi, K. Bakhtiyari, J. C. Ju'nior, "An Intrusion Detection and Prevention System in Cloud Computing: A Systematic Overview", *Journal of Network and Computer Applications*, vol. 36, pp. 25-41, 2013.

BIOGRAPHY

Namish Khanna pursuing Bachelor of Engineering from Chandigarh University in Computer Science and engineering (2017-2021). Completed Higher Secondary Education in 2017- and Six-Weeks Training in Machine Learning.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 5, May 2020||



Harnoor Singh pursuing Bachelor of Engineering from Chandigarh University in Computer Science and engineering (2017-2021). Completed Higher Secondary Education in 2017- and Six-Weeks Training in Machine Learning.

Aditya Sharma pursuing Bachelor of Engineering from Chandigarh University in Computer Science and engineering (2017-2021). Completed Higher Secondary Education in 2017- and Six-Weeks Training in Machine Learning.