Analysis of Machine learning algorithms with Feature Selection for Intrusion Detection

[1]Ramini Devi Priya, [2]Nithish Pallerla

[1]Information Technology,
[1]BV Raju Institute of Technology, City, Narsapur, India

***Abstract*— The most recent dataset, UNSW-NB15, is used to train machine learning classifiers. The dataset is trained using the chosen classifiers, including K-Nearest Neighbors (KNN), Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM) classifiers. The dataset is subjected to a filter-based feature selection technique to eliminate redundant and pointless characteristics. The machine learning classifiers are compared in this investigation. Comparative analysis of these machine learning classifiers is done, and the performance of classifiers is quantified in terms of Accuracy, Mean Squared Error (MSE), Precision, Recall, True Positive Rate (TPR), and False Positive Rate (FPR).**

***Index Terms*—Feature Selection, Comparative Analysis, Intrusion Detection.**

I. INTRODUCTION (HEADING 1)

An intrusion detection system (IDS) is a hardware or software program that keeps an eye out for hostile activities or policy breaches on a network or in a system. Every intrusion activity or violation is often recorded centrally using a security information and event management (SIEM) system, alerted to an administrator, or both. In order to separate harmful activity from false alerts, a SIEM system integrates outputs from many sources and employs alarm filtering mechanisms. Despite monitoring networks for potentially harmful behavior, intrusion detection systems are often prone to false alarms. Thus, enterprises must adjust their IDS products after initial installation. It entails correctly configuring intrusion detection systems to distinguish between legitimate network traffic and malicious activities.

The Intrusion detection can be done in the following 5 ways such as:

1. **Network Intrusion Detection System (NIDS):**

Network intrusion detection systems (NIDS) are set up at a planned point within the network to examine traffic from all devices on the network. It performs an observation of passing traffic on the entire subnet and matches the traffic that is passed on the subnets to the collection of known attacks. Once an attack is identified or abnormal behavior is observed, the alert can be sent to the administrator. An example of an NIDS is installing it on the subnet where firewalls are located in order to see if someone is trying crack the firewall.

2. **Host Intrusion Detection System (HIDS):**

Host intrusion detection systems (HIDS) run on independent hosts or devices on the network. A HIDS monitors the incoming and outgoing packets from the device only and will alert the administrator if suspicious or malicious activity is detected. It takes a snapshot of existing system files and compares it with the previous snapshot. If the analytical system files were edited or deleted, an alert is sent to the administrator to investigate. An example of HIDSusage can be seen on mission critical machines, which are not expected to change their layout.

3. **Protocol-based Intrusion Detection System (PIDS):**

Protocol-based intrusion detection system (PIDS) comprises of a system or agent that would consistently resides at the front end of a server, controlling and interpreting the protocol between a user/device and the server. It is trying to secure the web server by regularly monitoring the HTTPS protocol stream and accept the related HTTP protocol. As HTTPS is un- encrypted and before instantly entering its web presentation layer then this system would need to reside in this interface, between to use the HTTPS.

4. **Application Protocol-based Intrusion Detection System (APIDS):**

Application Protocol-based Intrusion Detection System (APIDS) is a system or agent thatgenerally resides within a group of servers. It identifies the intrusions by monitoring and interpreting the communication on application specific protocols. For example, this would monitor the SQL protocol explicit to the middleware as it transacts with the database in the web server.

5. **Hybrid Intrusion Detection System**:

Hybrid intrusion detection system is made by the combination of two or more approachesof the intrusion detection system. In the hybrid intrusion detection system, host agent or system data is combined with network information to develop a complete view of the network system. Hybrid intrusion detection system is more effective in comparison to the other intrusion detection system. Prelude is an example of Hybrid IDS.

II. DIFFERENT CLASSES OF ATTACKS:

The

- Denial of Service (DoS):
  An attacker tries to prevent legitimate users from using a service. For example, SYNflood, Smurf and teardrop.
- User to Root (U2R) :
  An attacker has local access to the victim machine and tries to gain super-user privilege. For example, buffer overflow attacks.
- Remote to Local (R2L):
  An attacker tries to gain access to victim machine without having an account on it. For example, password guessing attack.
- Probe :
  An attacker tries to gain information about the target host. For example, port-scan andping-sweep.

III. Experiments

Pattern Recognition Scheme for Distributed Denial of Service (DDoS) Attacks in Wireless

In three of the most used network topologies, this study defines unique attack techniques illustrating in [1] Distributed Denial of Service (DDoS) attacks against target nodes within wireless sensor networks. For the purpose of detecting attacks, a decentralized pattern recognition system based on graph neurons (GN) is suggested. The plan examines the network's internal traffic flow for DDoS attack trends. We assume that the attack patterns are influenced by both the energy levels in the present and the rates at which various target nodes use energy. In order to evaluate the success of our implementation, the effects of various pattern update rates on the pattern recognition accuracies for the three network topologies are presented in the conclusion.

Techniques used :

The Graph Neuron (GN) is an in-network, distributed, pattern recognition algorithm which can form an associative memory overlay on the physical sensor network by interconnecting sensor nodes in a graph-like structure called the GN array. The Graph Neuron (GN) as a light-weight pattern recognition application was used to detect DDoS attack patterns in WSNs.

Network Traffic Analysis and Intrusion Detection Using Packet Sniffer

The term "packet sniffer" refers to computer software from [2] that can intercept and record data traveling through a digital network or a portion of a network. By placing the NIC card in promiscuous mode, the sniffer is able to capture and subsequently decode these packets. Depending on the person's intentions when decoding the data, the information can be used in any way. One can sniff all or only a portion of the traffic coming from a single computer within the network, depending on the network's structure. To access traffic from other systems on the network, there are some ways to get around traffic narrowing by switches. Consideration is given to packet sniffer functionality, Linux platform development, and intrusion detection applications.

Techniques used:

Address Resolution Protocol (ARP) is used for packet sniffer has been designed as a solution to many network problems. Sniffers are very hard to detect due to its passiveness but there is always a way, and some of them.

IV. Performance Metrics

- True Positive Rate (TPR):
  It is calculated as the ratio between the number of correctly predicted attacks and the total number of attacks. If all intrusions are detected then the TPR is 1 which is extremely rare for an IDS. TPR is also called Detection Rate(DR) or the Sensitivity.

- False Positive Rate (FPR):
  It is calculated as the ratio between the number of normal instances incorrectly classified as an attack and the total number of normal instances.

- False Negative Rate (FNR):
  False Negative means when a detector fails to identify an anomaly and classifies it as normal.

- Classification Rate (CR) or Accuracy:
  The CR measures how accurate the IDS is in detecting normal or anomalous traffic behavior. It is described as the percentage of all those correctly predicted instances to all the instances.

$$\text{Accuracy} = \frac{\text{True positives + False negatives}}{\text{Total number of samples}}$$

The performance of the model is done using this accuracy and this determines the proper working model of the algorithms.

V. Testing And Validation

Testing is done to look for mistakes. Testing is the process of looking for any flaws or weaknesses in a piece of work. It offers a

a method for evaluating the functionality of individual parts, subassemblies, assemblies, and/or a finished product. It is the process of testing software to make sure that it satisfies user expectations and meets requirements without failing in an unacceptable way. Several test types exist. Every test type responds to a certain testing requirement.

The proposed work is implemented in Python 3.6.4 with libraries scikitlearn, pandas, matplotlib and other mandatory libraries. The training dataset is UNSW-NB-15. Machine learning algorithm is applied such as decision tree, Logistic regression and Random forest. We used these machine learning algorithm and indentified intrusion. The result shows that intrusion detection is efficient using Random Forestalgorithm. Random forest achieves 100% accuracy, Naïve Baye's achieves around 91% accuracy, Logistic Regression achieves 95% accuracy, SVM achieves 96% accuracy, KNN achieves 99% accuracy.

VALIDATION:

To validate a system, one must watch how it behaves. The verification and validation processes serve as a check to make sure that a phase's output is consistent with both its input and the system's overall needs. To make sure it is consistent with the desired output, this is done. If not, use specific mechanisms for repair to satisfy the standards.

VI. RESULTS

The above-conducted experiment results are to be noted down in a table that specifies the accuracy of each algorithm separately:

Table 1 EXPERIMENTAL ANALYSIS RESULTS

Algorithm	Accuracy(%)
Logistic regression	95.00
SVM	96..00
Random Forest Classifier	100.00
KNN	99.00
Nave Baye's Classifier	91.00

VII. SCREENSHOTS

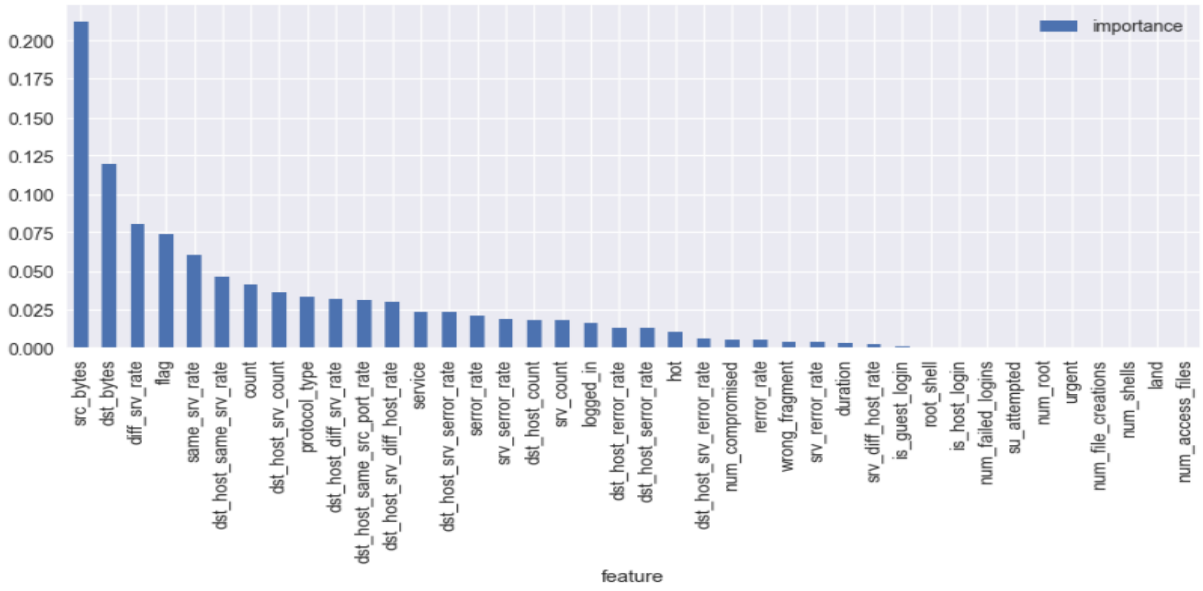

Fig.1. Importance Graph for Feature Selection

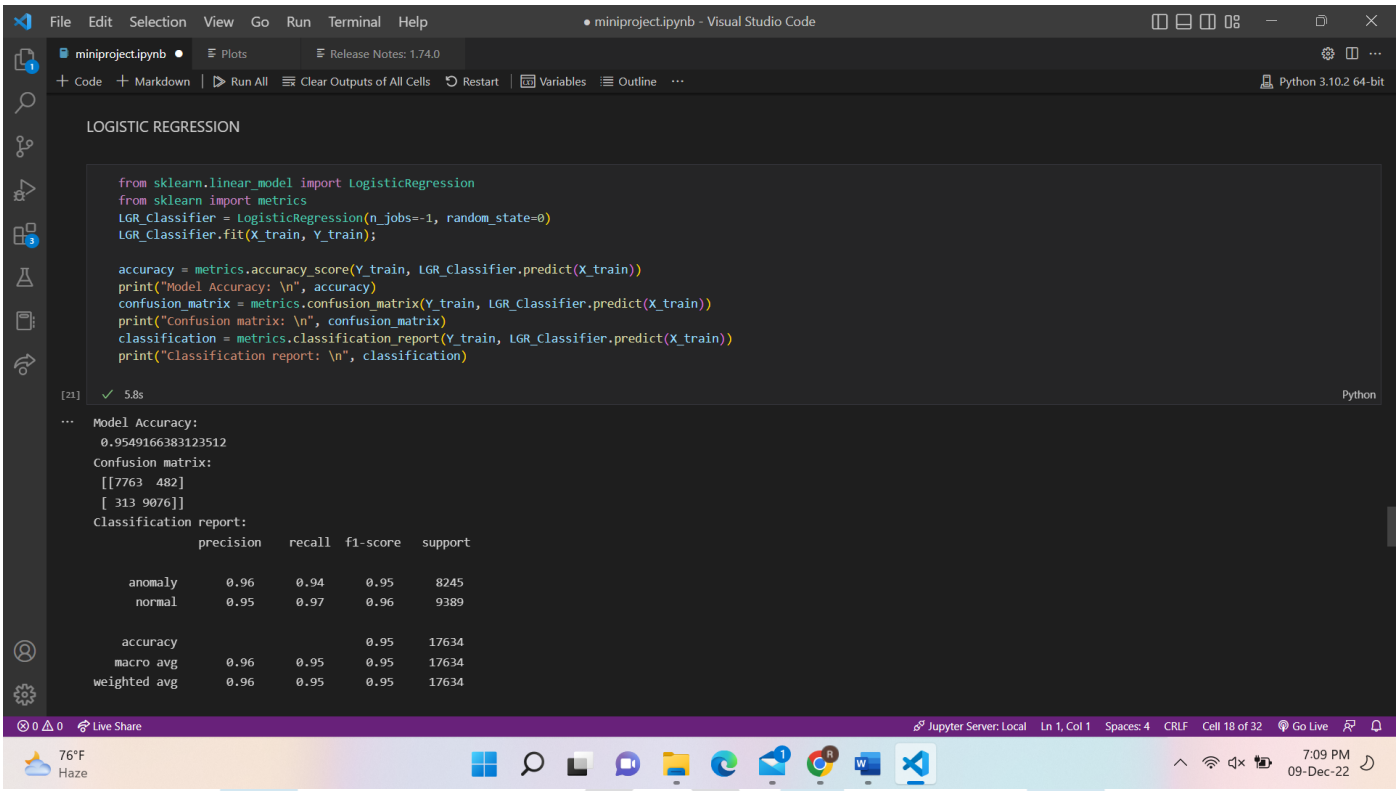

Fig 2. Output of Logistic Regression Algorithm

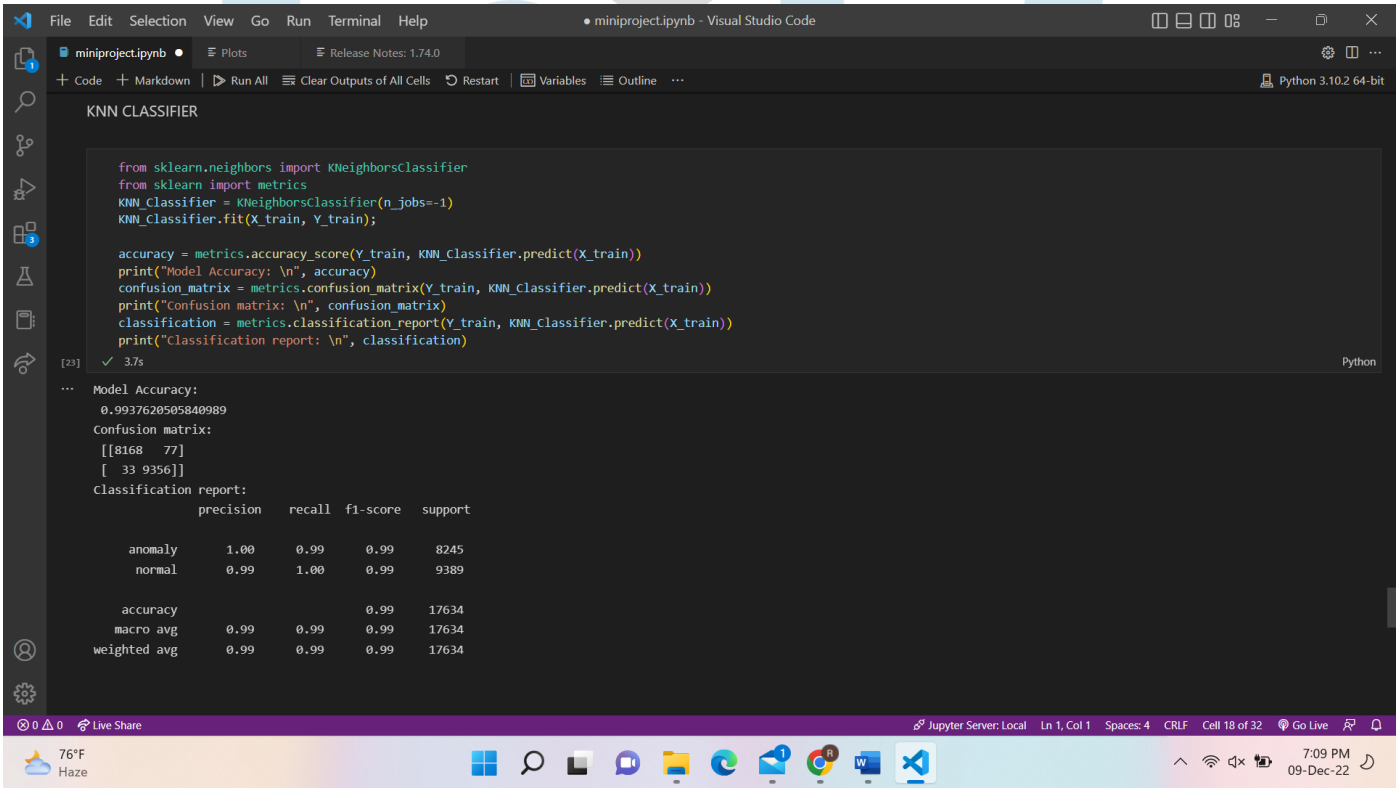

Fig 3: Output of KNN Algorithm

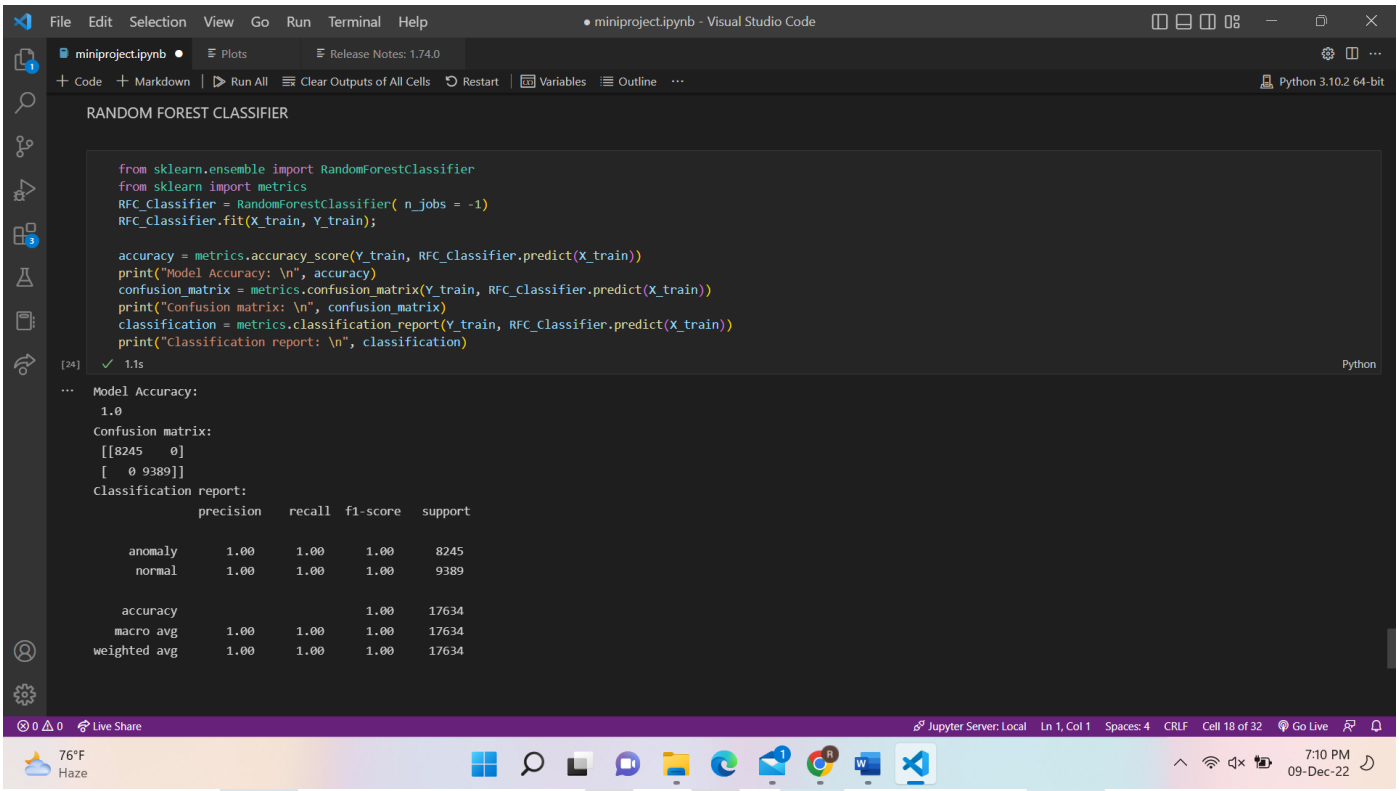

FIG 4: OUTPUT OF RANDOM FOREST ALGORITHM

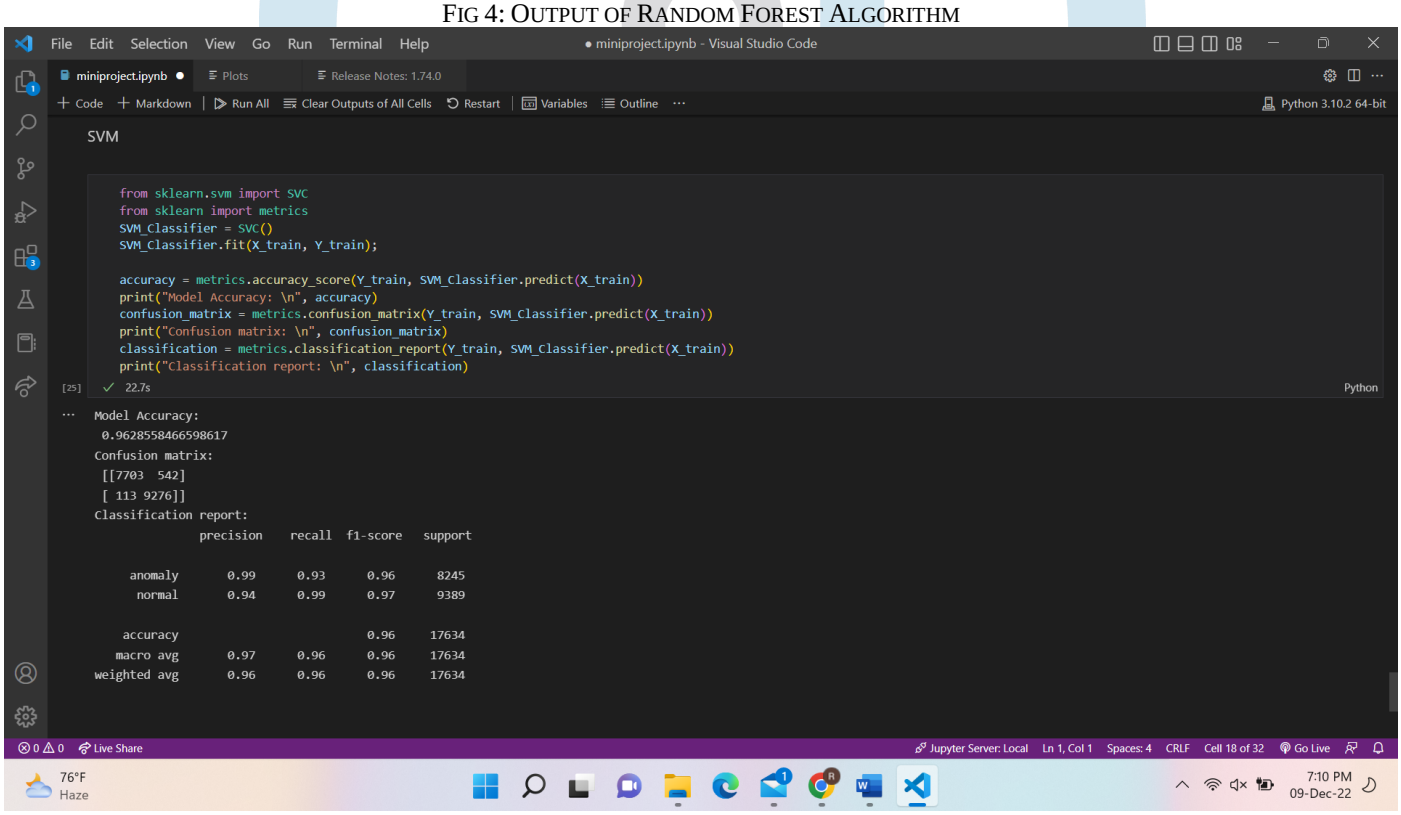

FIG 5: OUTPUT OF SVM ALGORITHM

© 2023 IJRTI | Volume 8, Issue 3 | ISSN: 2456-3315

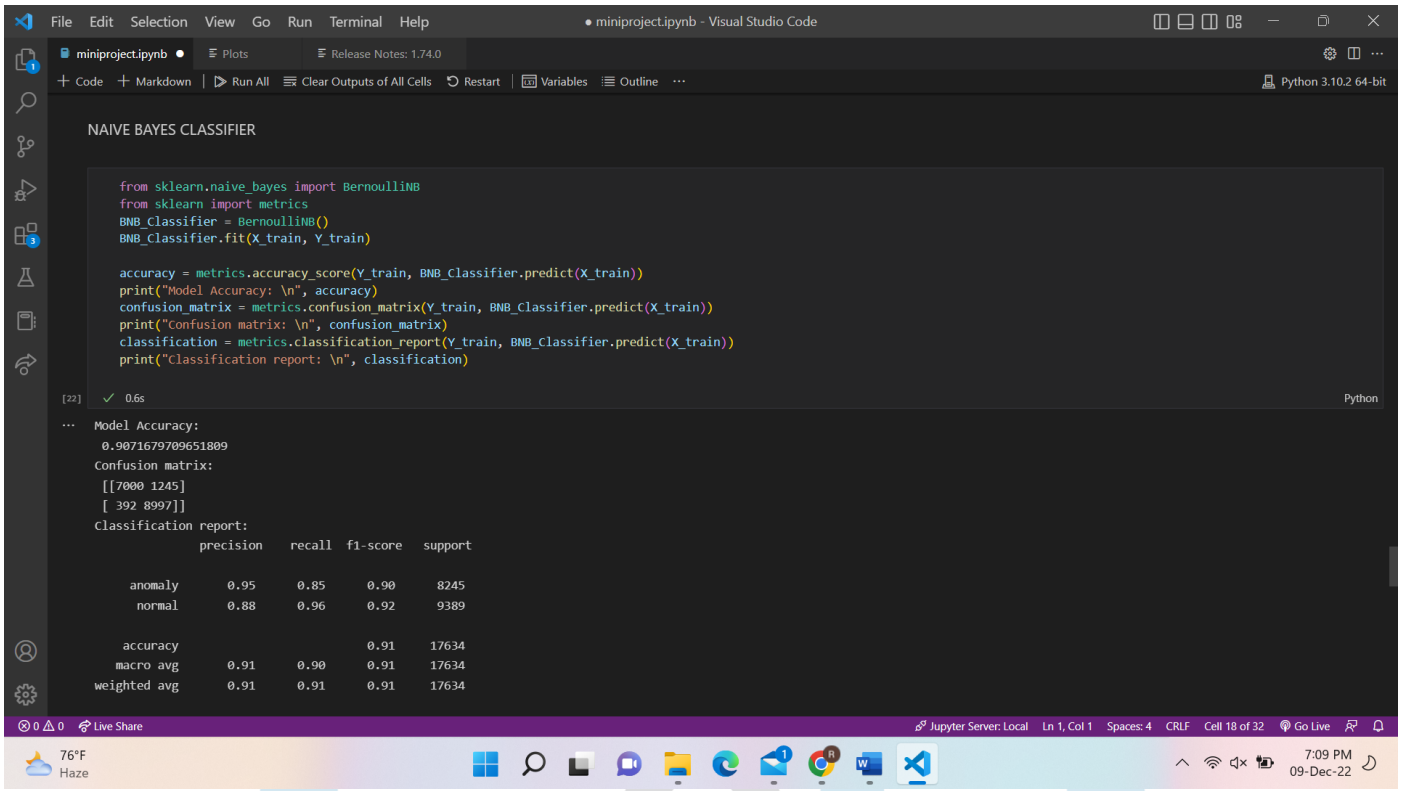

FIG 6: OUTPUT OF NAÏVE BAYES ALGORITHM

VIII. CONCLUSION

The main aim of Intrusion Detection System is to detect the attacks and malicious activities that occur within a network and to reduce the rate of false positives. By using the machine learning algorithms, the output of the IDS would be accurate, advanced and reliable. This system also shows the accuracy rate of the attacks that have been detected by the different machine learning algorithms that have been implemented. The incremental increase in the use of technology has led to huge amount of data that needs to be processed and stored securely for the users. Security is a major aspect for any user. If a system is secure, we can highly ensure user's privacy is high. The more secure the system, the more reliable it is. If an Intrusion Detection System is capable of providing good security for user's data, we can say that the developed Intrusion Detection System is good.

Random forest algorithm is used resulting in an accuracy of 100%.

IX. FUTURE WORK

The system that has been proposed can be made more reliable and efficient by applying more machine learning algorithms along with the ones that already have been applied so that infiltration can be detected quickly. To identify more attacks and offer greater protection and dependability, different attack types can also be categorized as intrusion classes. Hence, additional system development can contribute to raising the detection rate and decreasing false positive rates.

X. ACKNOWLEDGMENT

XI. The following individuals are thanked by the authors for their assistance in finishing our big project's investment. We would like to thank Prof. DR. K. Dasaradha Ramaiah, M.Tech. (Ph.D.), Professor and Head of the Department, Department of Information Technology, Internal Guide, for their inspiring assistance in completing the Mini Project. We acknowledge the assistance of the project coordinators, Ms. M. Krishna Prasanna, an assistant professor with a master's in technology, and Mr. M. Amarender Reddy, an assistant professor with a master's in technology (with a doctorate). Finally, we want to express our gratitude to our parents and friends for their unwavering support.

REFERENCES

[1] Sensor Networks, Baig, Zubair & Baqer, M & Khan, Asad. (2006), 1050 - 1054. 10.1109/ICPR.2006.147
[2] Qadeer, Mohammed & Iqbal, Arshad & Zahid, Mohammad & Siddiqui, Misbahur, Communication Software and Networks, International Conference on. 313-317. 10.1109/ICCSN.2010.104
[3] Kumar Vinod, Sangwan Prakash Om, "Signature Based Intrusion Detection System Using SNORT", IJCAIT, International Journal of Computer Applications & Information Technology, Vol. I, Issue III, November 2012.