**Supporting material:**

**Security Information and Event Management(SIEM) by**

**Azure Sentinel.**

***MA,***

***MA***

***University of ulster .***

**Table of contents:**

**1. Literature**

**2.Discussion of work.**

**3. Train and Testing (Verification and validation)**

**4. Work experience.**

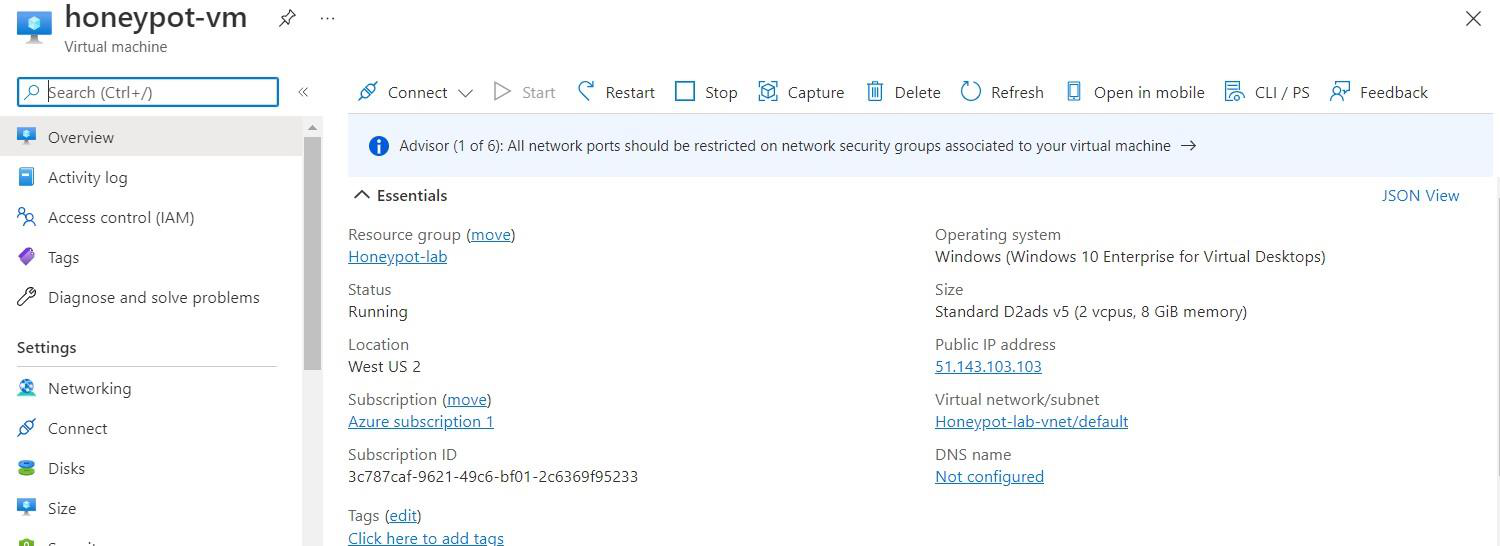
**Literature:**

The recent study was similar in terms of methodology and structure by ElTayeby [1] et al. The researcher used logs to filter and extract the insights from security log analysis than the security aspects of a system. Hence our research paper thoroughly studied their work on the log analysis which has some related work in the form of mapping as well. Case study has illustrated about how to pre-process the web logs and event logs and their extraction of raw data. However, mapping study was introduces in the year of 2010 but it only covers the web usage technique but our study is based on event viewing of failed attempts. Therefore our complete research relay upon cloud based system called as Microsoft Azure, Security tools techniques and Machine learning classifications methods.

Another research study by Agosti et al[2]. its work is related to our paper whereas it has mapping techniques over security log analysis but not on brute force attacks. However, their work is based on web engines log analysis and digital library log analysis, our main focus on mapping and investigation upon attackers.

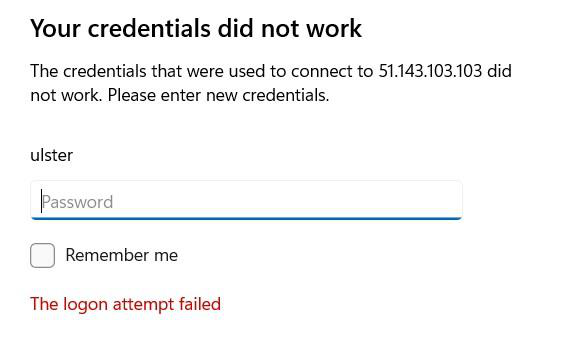
Lastly, our main study is related to outsider attacks which may be consider as DDos attacks .adversaries has done an extensive scan on publics IP addresses that scan been triggering on all the users denials in their firewall logs .Unexpectedly, shortly after Azure Sentinel was implemented, a suspicious attack was conducted against the company, resulting in massive amounts of unexpected log volumes. As Some users in Azure Sentinel have been subjected to suspicious attacks, and mitigating procedures have been developed to mitigate the damage. However, suspicious hackers execute a wide-ranging scan from a large number of remote servers against the customers public Ip addresses. While the scan may activate the user firewall by generating many rejections in the logs, the firewall logs are delivered to the Azure Sentinel (SIEM), resulting in a considerable rise in log ingestions and cost increases in Azure.

Description of work :Starting with a free trial account in Microsoft Azure that allows us to access the tools and technologies, numerous products were employed in this research procedure, including Virtual Machines, log analytics workspace, Microsoft defender cloud, and Microsoft Sentinel. First, make a new virtual machine resource called (Honeypot-lab). Virtual machines are typically used to host programmes and manage the common activities associated with servers. Furthermore, we will test the unsuccessful log on activities by establishing the Windows OS environment. Second, using the same approach, look for log analytics and construct a workspace by configuring all of the regions and advanced features connected with VMs.



The next step is to enable Microsoft Cloud Defender, which monitors security posture, hardens resources, and defends against cyber-attacks. Defend, secure, and continuously assess are the three key functions that Defender for Cloud plays in managing workloads and resources. Defender, on the other hand, secures resources by default and assists in pushing logs from VM to analytics workspace. Finally, Microsoft Sentinel, a SIEM product, provides intelligent security analytics, threat intelligence, and a single solution for attack detection, proactive hunting, and entity behaviour.

Now next part is to copying the honeypot -VM (virtual machine) IP address (51.143.103.103), pasting it into Remote Desktop, and pressing the connect button. The honeypot then connects to the remote environment automatically. However, before connecting to the desktop, it will ask for a username and password, which will be generated on the virtual machine, but it is recommended that you try failing to log in for testing purposes. For example, I generated a project login and a random password that are both valid, but I supplied the incorrect username 'ulster' for the testing audit failure purpose. It generates the audit failure in the security events almost instantly. The log on an unsuccessful attempt is shown in the diagram below.



As a result, the credentials we used for the test bed did work. Later, a remote desktop can be accessed by logging in with appropriate authentication. The next step is to open Event Viewer, which scans and aggregates text log files and functions as a database reporting programmer. However, our goal is to evaluate security events, which is also known as 'audits,' which reveals the action of security outcomes, which can be either unsuccessful or successful depending on whether the user or adversaries tries to get inside.

All audits are being uploaded to Windows PowerShell ISE with the use of a third-party (IP Geolocation) API key to gather all of the attackers' or users' logs and geolocations. However, because the Windows firewall protects audits from being sent to third-party apps, it must be turned off in order to obtain location information. The next step is to open ISE and create a new script,

[https://github.com/joshmadakor1/Sentinel-](https://github.com/joshmadakor1/Sentinel-Lab/blob/ef4cb54f7068715ca208c13e2a7c0a43ac58b915/Custom_Security_Log_Exporter.ps1)

[Lab/blob/ef4cb54f7068715ca208c13e2a7c0a43ac58b915/Custom\_Security\_Log\_Exporter.ps1](https://github.com/joshmadakor1/Sentinel-Lab/blob/ef4cb54f7068715ca208c13e2a7c0a43ac58b915/Custom_Security_Log_Exporter.ps1)

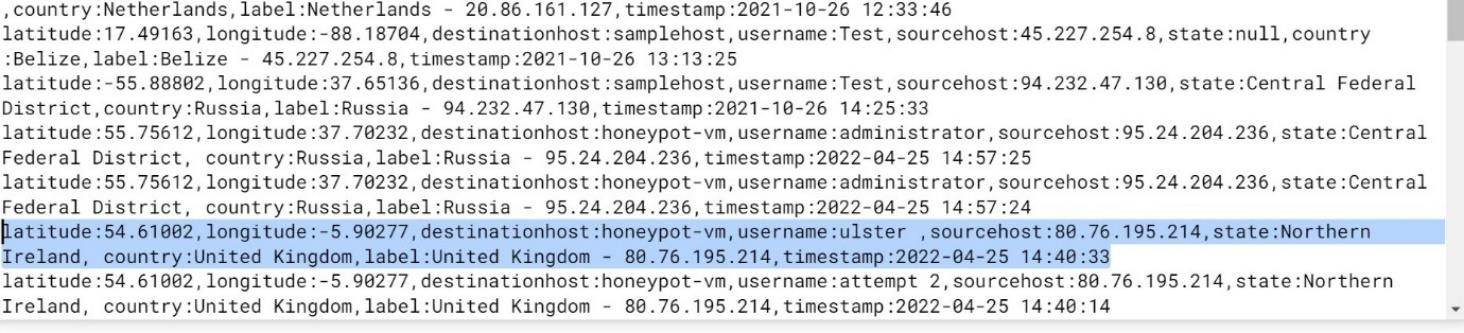
[(following hyperlink directs to the PowerShell script which is Geolocation)](https://github.com/joshmadakor1/Sentinel-Lab/blob/ef4cb54f7068715ca208c13e2a7c0a43ac58b915/Custom_Security_Log_Exporter.ps1)

paste the resources copied from GITHUB, then retrieve the personal API key issued by the IP Geolocation

application and run the script. ISE gathers all audit failure information for security events, then divides the live

data into a text file called "Failed RDP." The following figure tells the collection of script sample which is a test

bed .here IP is different because I have use many free trail accounts for experiencing more knowledge.



The Extract feature in the workspace we've established is trained using the example log files in the unsuccessful

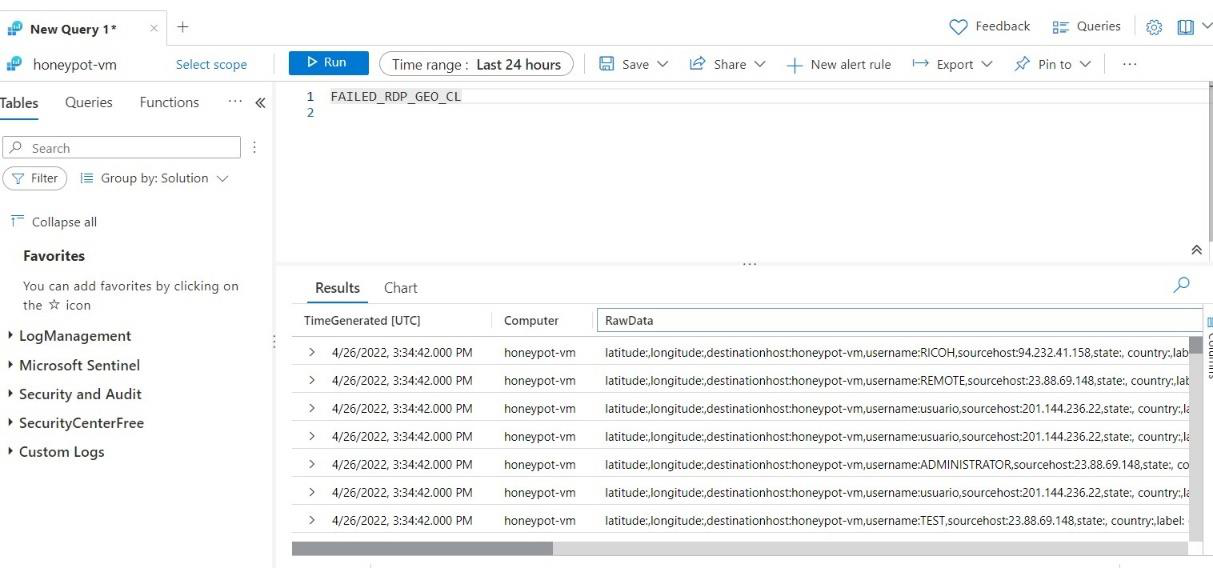
RDP log. However, due to a failed logon attempt, I addressed the geolocation of the erroneous username (ulster)

credentials, which are provided below with latitude, longitude, username, source host (Ip address), and country.

"latitude: 54.61002, longitude: -5.90277, destination host: honeypot-vm, username: ulster,

sourcehost:51.143.103.103, state: Northern Ireland, country: United Kingdom, timestamp:2022-04-25 14:40:33"

.



The preceding image depicts the result of raw data that was ingested after running the query 'FAILED RDP GEO'.

It has documented the audit's accomplishments and shortcomings from beginning to end. Workspace, on the other

hand, collects raw data rather than extracted data that is broken down into latitude, longitude, source host, nation,

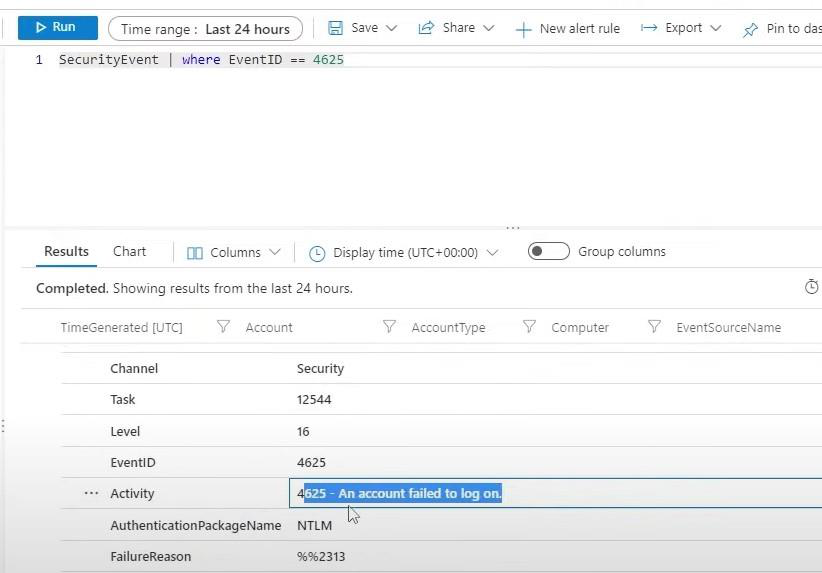
timestamp, and label.

Based on the events we have specific ID generated i.e., 4625 – account failed log on, 4634- when account logged

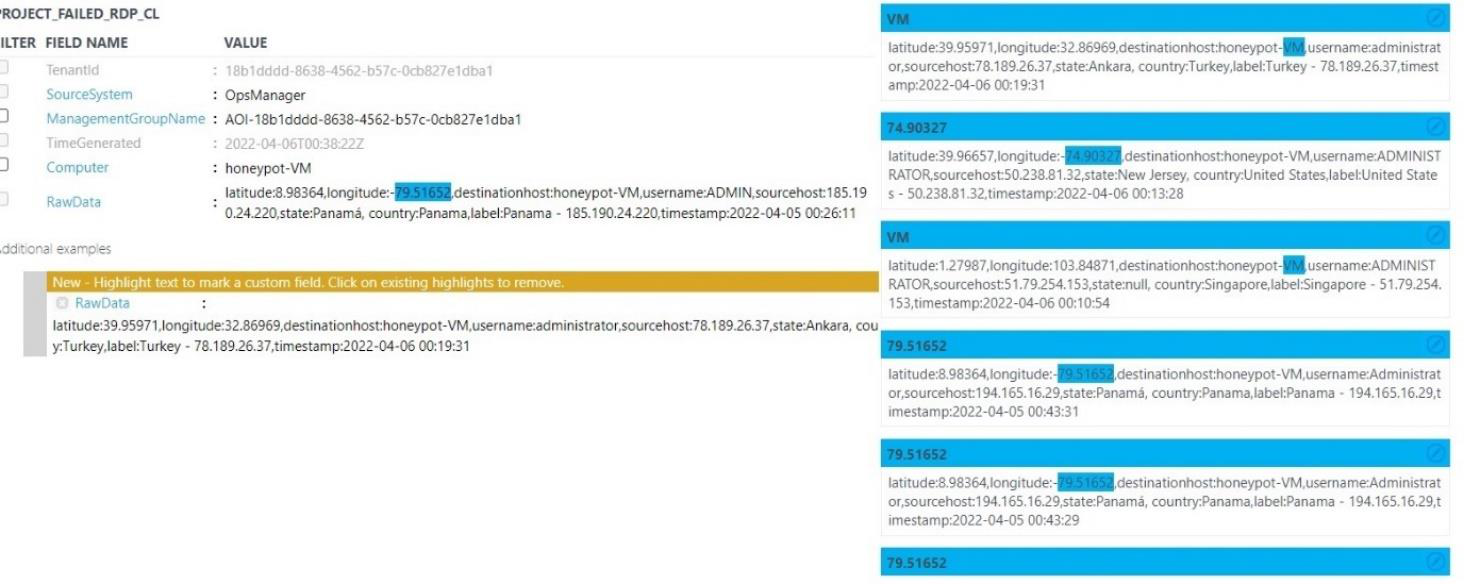
off , 8002 – a process allowed to run.

To check security events id query has to run as shoed below

‘ Security Events | where Event ID == 4625 ‘



Create a custom field for log analytics to determine what type of data value should be displayed. The second step is to double-check the exact field that has to be removed. For example, in the figure below, longitude is extracted by validating if all of the data relates to longitude or not. If any of the extracts are incorrect at first, they must be corrected and edited using modification.



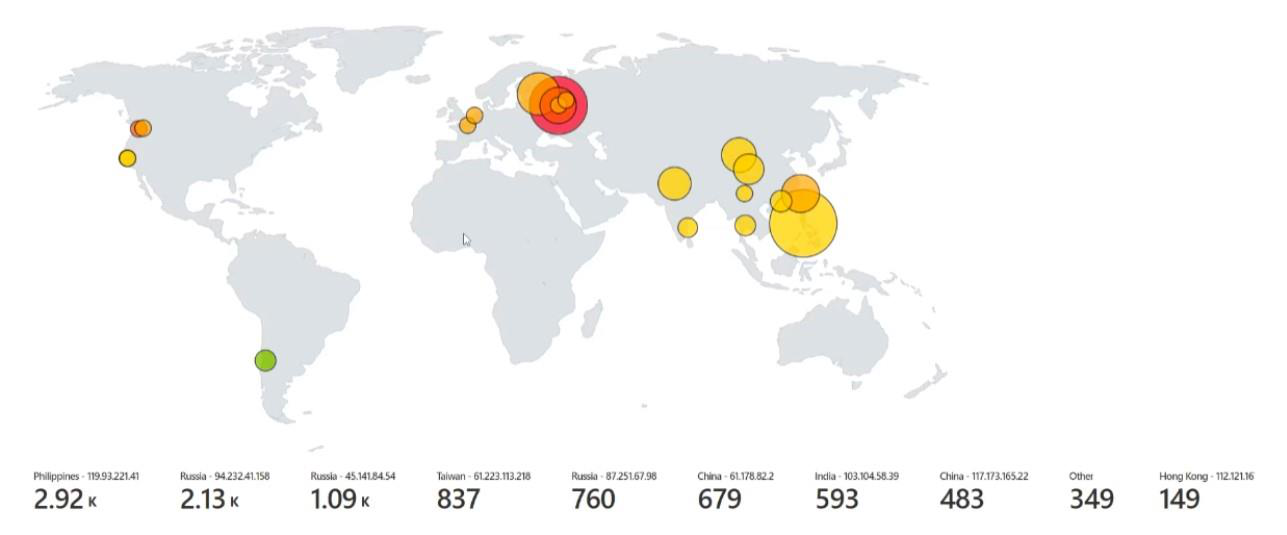
Overall, Custom fields accept only collected data when records satisfy the defined criteria and are added to the Analytics workspace after field extraction. Only live log data can be monitored or saved after fields have been extracted.

From now I have divided my work into two parts, investigation, and prediction. Moreover, by investigation I’ll try to find the vulnerabilities, attackers next movement and to prevent the organization from future attacks.

Prediction analysis helps to predict the future outcomes based on that can stop pre-attacks

Strong investigation has been made in Hunting by sentinel, which uses queries to proactively identify the behaviour of dangers and entities. Before that I have used sentinel mapping for better investigation. Sentinel offers a workbook based on geodata findings that enables for exact display after constructing the query.

“ FAILED\_ RDP\_ GEO\_CL | count event count=count() by sourcehost CF, latitude CF, longitude CF, country CF, label CF, destinationhost CF | where destinationhost |CF!= "samplehost" “

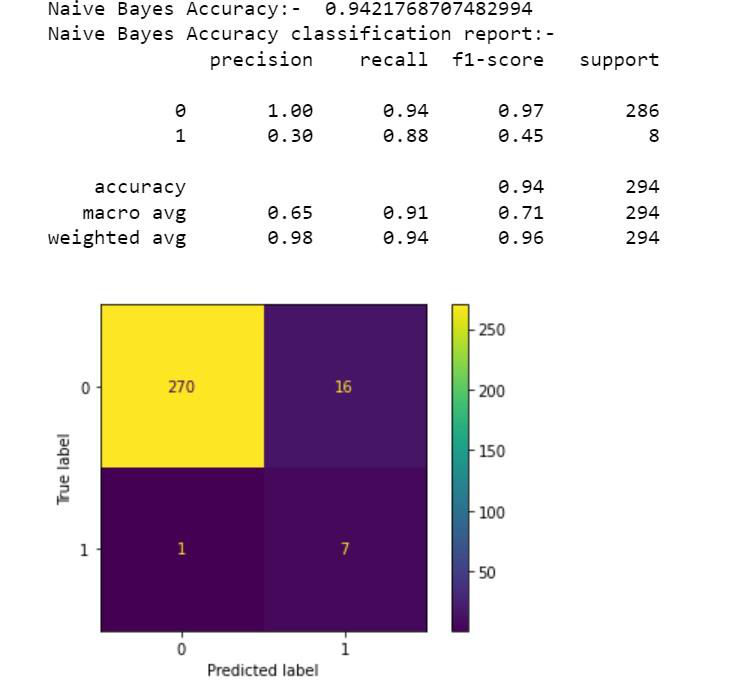


So, after manually refreshing the Power Shell ISE, data has been instantly delivered into the Sentinel mapping display, allowing all the attackers' co-ordinates from all over the world. so that they can be easily caught Each occurrence occurs in a distinct area with different IP addresses, according to 2 hours of live attacks mapping. Russia has a higher amount of log events, over 5,000, which are tracked using three distinct IP addresses: 2.13k (Events) – 94.232.41.158(IP), 1.09k (Events) – 45.141.84.54(IP), and 760(Events) – 87.251.67.98. (IP). However, Hong Kong has the fewest number of incidents, indicating that it has fewer rivals. To determine whether the suspicious attacker is an insider or an external threat, I opted to detect the suspicious attacker's geo location. By contemplating my test bed, which falls under the category of insider, I imply that an inside user of an organisation needs access to sensitive company information.

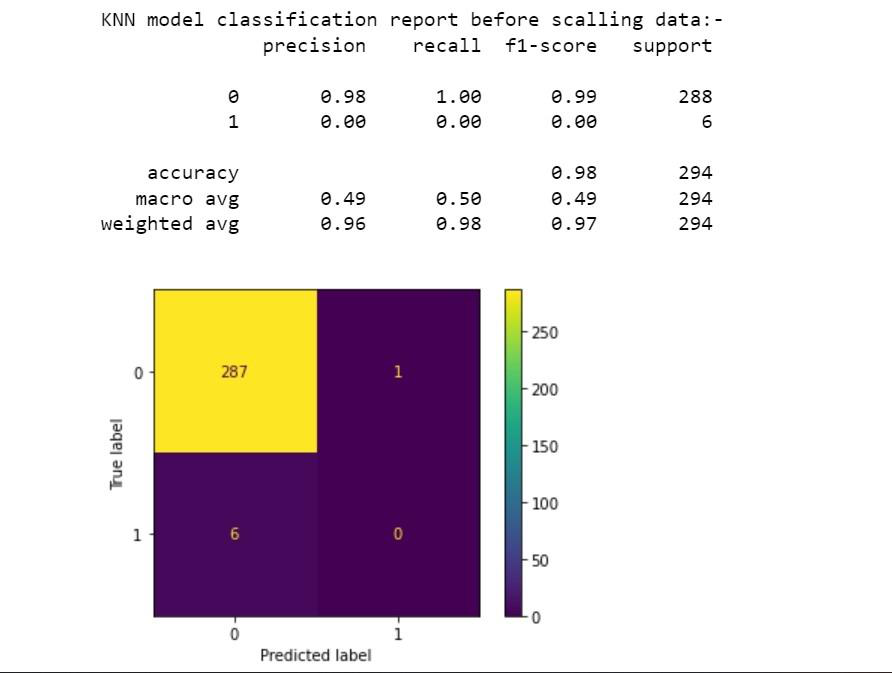
from defence to hacker To protect all of the risks and vulnerabilities, I switched places. Took part in the investigation as a hunter, hunting the initial access and then conducting query methods to prove that the adversaries were attempting to gain a footing within a network. Furthermore, Accounts and Host are the entities under examination, while T1078 is the mechanism utilised by attackers to create legal accounts. Enforce the usage of strong passwords and do not reuse them across numerous resources and services to prevent dangers.

**Validation and testing:**

For predictions used a few approaches such as KNN and Nave Bayes to capture all of the normalised data produced into a CSV file in order to perform better prediction analysis.



Above image illustrates that naïve bayes method, after training the model accuracy obtained as 94 percent .



I used Nave Bayes, SVM (Support Vector Machine), and KNN (K-nearest neighbour) for training and testing our prediction comparative analysis. From all the above mentioned, KNN is the best method for precision, resulting in a 98 percent accuracy. However, Nave Bayes has a score of 94, and SVM has a score of 96. Overall, the final mark of 21 might be attackers out of 900 users, based on train and test and test bed outcomes. In 21, on the other hand, I'm one of the suspicious attackers who attempted to gain access to the organisation.

Because of prediction I choose only three methods. Hence out of three only KNN gives the best accuracy about 98 percent. Some active security advice, on the other hand, help prevent future assaults. To begin, the machines should encrypt temporary discs, caches, and data transferred between storage and processing. Second, all network ports should be restricted to a single network security group. Finally, a vulnerability assessment solution should be installed on all internal machines.

**Work Experience:**

During the process of work, I have gone through all the investigations thoroughly, how log events occurs, transit of data, ingestion of raw data from virtual machine to cloud. Especially I put my hands on cloud servers which helps to gain more knowledge on SIEM. Working experience with cloud is good part because it contains all the servers, technology, tools, and AI, it has prebuilt query languages which makes our work easier and cost budget. In some way found difficulties managing with Azure free trail accounts but it makes interesting. Based on the limited resources on having public IP addresses and licenses, I had done a limited process in progress.

References :   
[1] .On Vulnerability and Security Log analysis: A Systematic Literature Review on Recent Trends   
 Omar ElTayeby and Wenwen Dou. 2016. A Survey on Interaction Log Analysis for Evaluating Exploratory Visualizations. In Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV ’16). Association for Computing Mach   
[2] . Tasawar Hussain, Sohail Asghar, and Nayyer Masood. 2010. Web usage mining: A survey on preprocessing of web log file. 2010 International Conference on Information and Emerging Technologies (2010).

[https://doi.org/10.1109/iciet.2010. 5625730](https://doi.org/10.1109/iciet.2010.%205625730)   
[[3] .](https://doi.org/10.1109/iciet.2010.%205625730)  [EDoS Attack in Azure Sentinel: Detect & Mitigate | Managed Sentinel](https://www.managedsentinel.com/edos-attack-azure-sentinel/)

**Additional code :**

I have made progress of training and testing of KNN, SVM and NB for the following code.



failed rd.ipynb

Another analysis code for overall profile reviewing.



Perfect.ipynb