**NumPy** is a Python library used for **working with arrays**. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

 Pandas is **a Python library used for working with data sets**.

• DOS (Deni of service): a denial-of-service attack is a type of attack in which the hacker generates computing or memory resources that are too busy or too saturated to meet legitimate network demands, thus preventing users from accessing memory resources.

• Probe: its actions are not really attacks since they are not destructive, they do not prevent an entity from functioning properly, but allow to acquire information sometimes crucial to conduct a larger attack later.

• U2R (User to Root attacks): a remote user attack is an attack in which a user sends packets to a machine via the Internet, which it does not have access to in case exposing the vulnerabilities of the machine and exploiting the privileges that a local user would have on the machine.

• R2L (Remote to Local access): its attacks are operations in which the hacker starts on the system with a normal user account and tries to abuse system vulnerabilities in order to obtain super user rights.

For more about data you can visit to <http://www.unb.ca/cic/datasets/nsl.html>

#### **BASIC FEATURES OF EACH NETWORK CONNECTION VECTOR**

1. **Duration:**Length of time duration of the connection
2. **Protocol\_type:**Protocol used in the connection
3. **Service:**Destination network service used
4. **Flag:**Status of the connection – Normal or Error
5. **Src\_bytes:**Number of data bytes transferred from source to destination in single connection
6. **Dst\_bytes:**Number of data bytes transferred from destination to source in single connection
7. **Land:**if source and destination IP addresses and port numbers are equal then, this variable takes value 1 else 0
8. **Wrong\_fragment:**Total number of wrong fragments in this connection
9. **Urgent:**Number of urgent packets in this connection. Urgent packets are packets with the urgent bit activated.
10. **Hot:**Number of „hot‟ indicators in the content such as: entering a system directory, creating programs and executing programs.
11. **Num\_failed \_logins:**Count of failed login attempts.
12. **Logged\_in Login Status**: 1 if successfully logged in; 0 otherwise.
13. **Num\_compromised**: Number of “compromised’ ‘ conditions.
14. **Root\_shell:**1 if root shell is obtained; 0 otherwise.
15. **Su\_attempted:**1 if “su root” command attempted or used; 0 otherwise.
16. **Num\_root:**Number of “root” accesses or number of operations performed as a root in the connection.
17. **Num\_file\_creations:**Number of file creation operations in the connection.
18. **Num\_shells:**Number of shell prompts.
19. **Num\_access\_files**: Number of operations on access control files .
20. **Num\_outbound\_cmds:**Number of outbound commands in an ftp session.
21. **Is\_hot\_login:**1 if the login belongs to the “hot” list i.e., root or admin; else 0.
22. **Is\_guest\_login:**1 if the login is a “guest” login; 0 otherwise .
23. **Count:**Number of connections to the same destination host as the current connection in the past two seconds
24. **Srv\_count:**Number of connections to the same service (port number) as the current connection in the past two seconds.
25. **Serror\_rate:**The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in count (23 )
26. **Srv\_serror\_rate:**The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in srv\_count (24)
27. **Rerror\_rate:**The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in count (23)
28. **Srv\_rerror\_rate:**The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in srv\_count (24)
29. **Same\_srv\_rate:**The percentage of connections that were to the same service, among the connections aggregated in count (23)
30. **Diff\_srv\_rate:**The percentage of connections that were to different services, among the connections aggregated in count (23)
31. **Srv\_diff\_host\_ rate:** The percentage of connections that were to different destination machines among the connections aggregated in srv\_count (24)
32. **Dst\_host\_count:** Number of connections having the same destination host IP address.
33. **Dst\_host\_srv\_ count:** Number of connections having the same port number.
34. **Dst\_host\_same \_srv\_rate:** The percentage of connections that were to the same service, among the connections aggregated in dst\_host\_count (32) .
35. **Dst\_host\_diff\_ srv\_rate:** The percentage of connections that were to different services, among the connections aggregated in dst\_host\_count (32)
36. **Dst\_host\_same \_src\_port\_rate:** The percentage of connections that were to the same source port, among the connections aggregated in dst\_host\_srv\_c ount (33) .
37. **Dst\_host\_srv\_ diff\_host\_rate:** The percentage of connections that were to different destination machines, among the connections aggregated in dst\_host\_srv\_count (33).
38. **Dst\_host\_serro r\_rate:** The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in dst\_host\_count (32).
39. **Dst\_host\_srv\_s error\_rate:** The percent of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in dst\_host\_srv\_c ount (33).
40. **Dst\_host\_rerro r\_rate:** The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in dst\_host\_count (32) .
41. **Dst\_host\_srv\_r error\_rate:** The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in dst\_host\_srv\_c ount (33).