***Introduction***

The basic idea of internet connection is as follows:

1. A connection is initiated from the port of a machine (the source) to a port on another or the same machine (the destination).
2. Information is transmitted along this connection in form of ***packets***. These packets may vary in size.
3. Sometimes, the packets may be very large, in which case they are broken down into ***fragments***or ***segments***via a process known as ***fragmentation***.
4. Packets may be sent either source-to-destination (i.e., forward direction) or destination-to-source (i.e., backward direction).
5. Among these packets, there is usually a special one that describes and allows for authentication of all the other packets. This special packet is known as the ***header***.
6. During or at the end of packet transmission in either the forward or backward direction, the status of the connection or packet receipt is designated via a set of ***flags***. They include: *FIN, PSH, WAIT, SYN, REST, and ACK*. A flag might be raised depending on the status of the packets or connection.

***Features***

The features of interest in the dataset are outlined below:

1. Src\_Port, - Originating port for connection
2. Dst\_Port, - Destination port for connection
3. Protocol, - Connection protocol (HTTP, HTTPS, UDP, TCP etcetera)
4. Flow\_Duration, - Duration of connection
5. Tot\_Fwd\_Pkts, - Total number of packets transmitted forward over connection lifetime
6. Tot\_Bwd\_Pkts, - Total number of packets transmitted backward over connection lifetime
7. TotLen\_Fwd\_Pkts, - Total size of packets transmitted forward over connection lifetime
8. TotLen\_Bwd\_Pkts, - Total size of packets transmitted backward over connection lifetime
9. Fwd\_Pkt\_Len\_Max, - Size of largest packet transmitted forward over connection lifetime
10. Fwd\_Pkt\_Len\_Min, - Size of smallest packet transmitted forward over connection lifetime
11. Fwd\_Pkt\_Len\_Mean, - Mean size of packets transmitted forward over connection lifetime
12. Fwd\_Pkt\_Len\_Std, - Standard deviation of size of packets transmitted forward over connection lifetime
13. Bwd\_Pkt\_Len\_Max, - Size of largest packet transmitted backward over connection lifetime
14. Bwd\_Pkt\_Len\_Min, - Size of smallest packet transmitted backward over connection lifetime
15. Bwd\_Pkt\_Len\_Mean, - Mean size of packets transmitted backward over connection lifetime
16. Bwd\_Pkt\_Len\_Std, - Standard deviation of size of packets transmitted backward over connection lifetime
17. Flow\_Byts/s, - Overall connection speed in bytes per second
18. Flow\_Pkts/s, - Overall connection speed in number of packets per second
19. Flow\_IAT\_Mean,
20. Flow\_IAT\_Std,
21. Flow\_IAT\_Max,
22. Flow\_IAT\_Min,
23. Fwd\_IAT\_Tot,
24. Fwd\_IAT\_Mean,
25. Fwd\_IAT\_Std,
26. Fwd\_IAT\_Max,
27. Fwd\_IAT\_Min,
28. Bwd\_IAT\_Tot,
29. Bwd\_IAT\_Mean,
30. Bwd\_IAT\_Std,
31. Bwd\_IAT\_Max,
32. Bwd\_IAT\_Min,
33. Bwd\_PSH\_Flags, - PSH flag raised during backward transmission?
34. Fwd\_Header\_Len, - Size of header for forward packets.
35. Bwd\_Header\_Len, - Size of header for backward packets.
36. Fwd\_Pkts/s, - Forward connection speed in number of packets per second
37. Bwd\_Pkts/s, - Backward connection speed in number of packets per second
38. Pkt\_Len\_Min, - Size of smallest packet transmitted over connection lifetime
39. Pkt\_Len\_Max, - Size of largest packet transmitted over connection lifetime
40. Pkt\_Len\_Mean, - Mean packet size over connection lifetime
41. Pkt\_Len\_Std, - Standard deviation of packet sizes over connection lifetime
42. Pkt\_Len\_Var, - Variance of packet sizes over connection lifetime
43. FIN\_Flag\_Cnt, - Number of *FIN* flags raised over connection lifetime.
44. SYN\_Flag\_Cnt, - Number of *SYN* flags raised over connection lifetime.
45. RST\_Flag\_Cnt, - Number of *RST* flags raised over connection lifetime.
46. PSH\_Flag\_Cnt, - Number of *PSH* flags raised over connection lifetime.
47. ACK\_Flag\_Cnt, - Number of *ACK* flags raised over connection lifetime.
48. Down/Up\_Ratio, - Ratio of
49. Pkt\_Size\_Avg, - Average packet size over entire connection lifetime
50. Fwd\_Seg\_Size\_Avg, - Average size of forward packet segments over connection lifetime
51. Bwd\_Seg\_Size\_Avg, - Average size of backward packet segments over connection lifetime
52. Subflow\_Fwd\_Pkts,
53. Subflow\_Fwd\_Byts,
54. Subflow\_Bwd\_Pkts,
55. Subflow\_Bwd\_Byts,
56. Init\_Bwd\_Win\_Byts,
57. Fwd\_Act\_Data\_Pkts,
58. Active\_Mean, - Average packet/connection active time
59. Active\_Std, - Standard deviation of packet/connection active time
60. Active\_Max, - Maximum packet/connection active time
61. Active\_Min, - Minimum packet/connection active time
62. Idle\_Mean, - Average packet/connection idle time
63. Idle\_Std, - Standard deviation of packet/connection idle time
64. Idle\_Max, - Maximum packet/connection idle time
65. Idle\_Min, - Minimum packet/connection idle time