Cyber-adversaries are becoming more sophisticated in their efforts to avoid detection, and many modern malware tools are already incorporating new ways to bypass antivirus and other threat detection measures. Software to detect network attacks protects a computer network from unauthorized users, including perhaps insiders. We need new methods to find out the attacks.

This is a multi-class classification problem. The given dataset contains 41 feature columns and 1 label column. The task is to classify the data into 5 types of network attacks, namely:-

ipsweep probe

back dos

satan probe

portsweep probe

normal

Note:- the label normal refers to data showing no attack.

Consider the following details about the feature columns:-

duration: length (number of seconds) of the connection

protocol_type: type of the protocol, e.g. tcp, udp, etc.

service: network service on the destination, e.g., http, telnet, etc.

src_bytes: number of data bytes from source to destination

dst_bytes: number of data bytes from destination to source

flag: normal or error status of the connection land: 1 if connection is from/to the same host/port; 0 otherwise

wrong_fragment: number of ``wrong'' fragments

urgent: number of urgent packets

hot: number of ``hot'' indicators

num_failed_logins: number of failed login attempts

logged_in: 1 if successfully logged in; 0 otherwise

num_compromised: number of ``compromised'' conditions

root_shell: 1 if root shell is obtained; 0 otherwise

su_attempted: 1 if ``su root'' command attempted; 0 otherwise

num_root: number of ``root'' accesses