Chosen Columns Based on Random Forest Computing Feature Importance:

Random Forest computes feature importance based on the concept of Gini Importance (also known as Mean Decrease in Impurity):

* At Each Split in a Tree: When a node is split using a feature, the impurity decrease is calculated (e.g., reduction in Gini impurity).
* Accumulated Over Trees: This decrease is accumulated for each feature over all trees in the forest.
* Normalized: The total decrease in impurity attributed to each feature is normalized to sum to 1.

Selecting Top N Features

* Compute Feature Importances: Use the trained Random Forest model to obtain importance scores for each feature.
* Rank Features: Sort the features based on their importance scores in descending order.
* Select Top N Features: Choose the top N features with the highest importance scores.

1. Avg. Bwd. Segment Size

2. Max\_Packet\_Length

3. Packet\_Length\_Std

4. Bwd\_Packet\_Length\_Max

5. Bwd\_Packet\_Length\_Std

6. Packet\_Length\_Variance

7. Average\_Packet\_Size

8. Bwd\_Packet\_Length\_Mean

9. Packet\_Length\_Mean

10. Total\_Length\_of\_Bwd\_Packets

11. Flow\_IAT\_Mean

12. Fwd\_IAT\_Mean

13. Subflow\_Bwd\_Bytes

14. Idle\_Mean

15. Fwd\_IAT\_Std

16. Idle\_Min

17. Destination\_Port

18. Flow\_IAT\_Std

19. Flow\_IAT\_Max

20. Init\_Win\_bytes\_backward

21. Attack

The following are a list of features included in the CIC-IDS2017 dataset:

**Feature**: Flow duration

* **Description**: Total duration of the network flow measured in microseconds, representing the time taken from the first to the last packet in the flow.

**Feature**: total Fwd Packet

* **Description**: Total number of packets transmitted in the forward direction (from source to destination) within the flow.

**Feature**: total Bwd packets

* **Description**: Total number of packets transmitted in the backward direction (from destination to source) within the flow.

**Feature**: total Length of Fwd Packet

* **Description**: Sum of the payload sizes of all packets transmitted in the forward direction, in bytes.

**Feature**: total Length of Bwd Packet

* **Description**: Sum of the payload sizes of all packets transmitted in the backward direction, in bytes.

**Feature**: Fwd Packet Length Min

* **Description**: Smallest packet size in bytes among all packets in the forward direction within the flow.

**Feature**: Fwd Packet Length Max

* **Description**: Largest packet size in bytes among all packets in the forward direction within the flow.

**Feature**: Fwd Packet Length Mean

* **Description**: Average packet size in bytes for all packets transmitted in the forward direction within the flow.

**Feature**: Fwd Packet Length Std

* **Description**: Standard deviation of packet sizes in bytes for all forward-direction packets, indicating the variability of packet size.

**Feature**: Bwd Packet Length Min

* **Description**: Smallest packet size in bytes among all packets in the backward direction within the flow.

**Feature**: Bwd Packet Length Max

* **Description**: Largest packet size in bytes among all packets in the backward direction within the flow.

**Feature**: Bwd Packet Length Mean

* **Description**: Average packet size in bytes for all packets transmitted in the backward direction within the flow.

**Feature**: Bwd Packet Length Std

* **Description**: Standard deviation of packet sizes in bytes for all backward-direction packets, reflecting the variability of packet size.

**Feature**: Flow Bytes/s

* **Description**: Average rate of bytes transmitted per second across the entire flow, calculated as the total bytes divided by the flow duration.

**Feature**: Flow Packets/s

* **Description**: Average rate of packets transmitted per second across the entire flow, calculated as the total packets divided by the flow duration.

**Feature**: Flow IAT Mean

* **Description**: Average inter-arrival time between packets within the flow, representing the average delay between successive packets.

**Feature**: Flow IAT Std

* **Description**: Standard deviation of the inter-arrival times between packets within the flow, showing the variability of packet intervals.

**Feature**: Flow IAT Max

* **Description**: Maximum inter-arrival time between packets in the flow, indicating the longest delay between two successive packets.

**Feature**: Flow IAT Min

* **Description**: Minimum inter-arrival time between packets in the flow, indicating the shortest delay between two successive packets.

**Feature**: Fwd IAT Min

* **Description**: Minimum inter-arrival time between packets in the forward direction, indicating the shortest delay between two forward packets.

**Feature**: Fwd IAT Max

* **Description**: Maximum inter-arrival time between packets in the forward direction, indicating the longest delay between two forward packets.

**Feature**: Fwd IAT Mean

* **Description**: Average inter-arrival time between packets in the forward direction, reflecting the mean delay between successive forward packets.

**Feature**: Fwd IAT Std

* **Description**: Standard deviation of inter-arrival times between forward-direction packets, showing the variability in delay between forward packets.

**Feature**: Fwd IAT Total

* **Description**: Total inter-arrival time for all packets in the forward direction, calculated as the sum of all inter-arrival times.

**Feature**: Bwd IAT Min

* **Description**: Minimum inter-arrival time between packets in the backward direction, indicating the shortest delay between two backward packets.

**Feature**: Bwd IAT Max

* **Description**: Maximum inter-arrival time between packets in the backward direction, indicating the longest delay between two backward packets.

**Feature**: Bwd IAT Mean

* **Description**: Average inter-arrival time between packets in the backward direction, reflecting the mean delay between successive backward packets.

**Feature**: Bwd IAT Std

* **Description**: Standard deviation of inter-arrival times between backward packets, showing the variability in delay between backward packets.

**Feature**: Bwd IAT Total

* **Description**: Total inter-arrival time for all packets in the backward direction, calculated as the sum of all inter-arrival times.

**Feature**: Fwd PSH flags

* **Description**: Count of PSH (Push) flags set in packets sent in the forward direction, often associated with TCP connections. Zero for UDP packets.

**Feature**: Bwd PSH Flags

* **Description**: Count of PSH (Push) flags set in packets sent in the backward direction, typically relevant in TCP connections. Zero for UDP packets.

**Feature**: Fwd URG Flags

* **Description**: Count of URG (Urgent) flags set in packets sent in the forward direction, indicating urgent data transmission for TCP connections. Zero for UDP packets.

**Feature**: Bwd URG Flags

* **Description**: Count of URG (Urgent) flags set in packets sent in the backward direction, indicating urgent data transmission for TCP connections. Zero for UDP packets.

**Feature**: Fwd Header Length

* **Description**: Total bytes used by packet headers in the forward direction, representing the overhead in packet transmission.

**Feature**: Bwd Header Length

* **Description**: Total bytes used by packet headers in the backward direction, representing the overhead in packet transmission.

**Feature**: FWD Packets/s

* **Description**: Rate of forward packets per second, calculated as the total forward packets divided by flow duration.

**Feature**: Bwd Packets/s

* **Description**: Rate of backward packets per second, calculated as the total backward packets divided by flow duration.

**Feature**: Packet Length Min

* **Description**: Minimum length in bytes of any packet in the flow, across both forward and backward directions.

**Feature**: Packet Length Max

* **Description**: Maximum length in bytes of any packet in the flow, across both forward and backward directions.

**Feature**: Packet Length Mean

* **Description**: Average length in bytes of packets within the flow, across both forward and backward directions.

**Feature**: Packet Length Std

* **Description**: Standard deviation of packet lengths within the flow, showing the variability in packet size across both directions.

**Feature**: Packet Length Variance

* **Description**: Variance in packet lengths within the flow, representing the spread of packet sizes around the mean length.

**Feature**: FIN Flag Count

* **Description**: Number of packets in the flow with the FIN flag set, indicating the sender's intention to terminate the connection in TCP-based protocols.

**Feature**: SYN Flag Count

* **Description**: Number of packets in the flow with the SYN flag set, which is used to initiate a connection in TCP-based protocols.

**Feature**: RST Flag Count

* **Description**: Number of packets in the flow with the RST (reset) flag set, typically used to abruptly close or reset a connection.

**Feature**: PSH Flag Count

* **Description**: Number of packets in the flow with the PSH (push) flag set, which instructs the receiver to push the data to the application as soon as possible.

**Feature**: ACK Flag Count

* **Description**: Number of packets in the flow with the ACK (acknowledgment) flag set, confirming receipt of packets in TCP-based protocols.

**Feature**: URG Flag Count

* **Description**: Number of packets in the flow with the URG (urgent) flag set, indicating urgent data in TCP-based protocols.

**Feature**: CWR Flag Count

* **Description**: Number of packets with the CWR (congestion window reduced) flag set, used to signal network congestion control adjustments.

**Feature**: ECE Flag Count

* **Description**: Number of packets with the ECE (ECN-Echo) flag set, which is part of explicit congestion notification for TCP to reduce network congestion.

**Feature**: down/Up Ratio

* **Description**: Ratio of download to upload bytes within the flow, providing a metric of the data direction balance.

**Feature**: Average Packet Size

* **Description**: Average size of packets in the flow, calculated as the total bytes divided by the total packets in both directions.

**Feature**: Fwd Segment Size Avg

* **Description**: Average size of segments in the forward direction, often associated with TCP segment sizes.

**Feature**: Bwd Segment Size Avg

* **Description**: Average size of segments in the backward direction, often associated with TCP segment sizes.

**Feature**: Fwd Bytes/Bulk Avg

* **Description**: Average number of bytes per bulk transfer in the forward direction, representing the data transfer rate in bursts.

**Feature**: Fwd Bulk Rate Avg

* **Description**: Average rate of bulk transfers in the forward direction, useful for analyzing data transfer patterns.

**Feature**: Bwd Bytes/Bulk Avg

* **Description**: Average number of bytes per bulk transfer in the backward direction, representing data transfer rate in bursts.

**Feature**: Bwd Bulk Rate Avg

* **Description**: Average rate of bulk transfers in the backward direction, useful for analyzing data transfer patterns.

**Feature**: Subflow Fwd Packets

* **Description**: Average number of packets per sub-flow in the forward direction, representing fragmented packet sequences within the flow.

**Feature**: Subflow Fwd Bytes

* **Description**: Average number of bytes per sub-flow in the forward direction, indicating byte distribution across sub-flows.

**Feature**: Subflow Bwd Packets

* **Description**: Average number of packets per sub-flow in the backward direction, representing fragmented packet sequences within the flow.

**Feature**: Subflow Bwd Bytes

* **Description**: Average number of bytes per sub-flow in the backward direction, indicating byte distribution across sub-flows.

**Feature**: Fwd Init Win bytes

* **Description**: Total bytes sent in the initial TCP window in the forward direction, often set at the start of a TCP connection.

**Feature**: Bwd Init Win bytes

* **Description**: Total bytes sent in the initial TCP window in the backward direction, part of the initial TCP connection handshake.

**Feature**: Fwd Act Data Pkts

* **Description**: Number of packets with at least 1 byte of TCP data payload in the forward direction, representing actively transmitted data packets.

**Feature**: Fwd Seg Size Min

* **Description**: Minimum observed size of a TCP segment in the forward direction, indicating the smallest segment used in data transmission.

**Feature**: Active Min

* **Description**: Minimum time duration a flow remains active before becoming idle, representing short-lived data exchanges.

**Feature**: Active Mean

* **Description**: Average time duration a flow remains active before becoming idle, showing the general active period of data transmission.

**Feature**: Active Max

* **Description**: Maximum time duration a flow remains active before becoming idle, indicating the longest continuous data transmission period.

**Feature**: Active Std

* **Description**: Standard deviation of active time durations, reflecting the variability in active periods within the flow.

**Feature**: Idle Min

* **Description**: Minimum time duration a flow remains idle before resuming activity, representing brief pauses in data transmission.

**Feature**: Idle Mean

* **Description**: Average time duration a flow remains idle, indicating the general pause duration in data transmission.

**Feature**: Idle Max

* **Description**: Maximum time duration a flow remains idle before resuming activity, showing the longest pause in data transmission.

**Feature**: Idle Std

* **Description**: Standard deviation of idle time durations, indicating variability in idle periods within the flow.