📊 Dataset Overview

NSL-KDD

- **Records:** 125,973
- **Classes:** 23 attack/normal labels
- **Attack Categories:** DoS, U2R, R2L, Probe
- **Features:** 41 traffic features + 1 label
    - Examples: duration, protocol type, service, flag, srcbytes,labelbytes,
- **Top 5 Frequent Attacks:** [to be listed based on counts]
- **Visualization:** Attack distribution chart

CICIDS2017

- **Records:** Large-scale real network traffic (millions of flows)
- **Attack Types:** DDoS, Brute Force, Web Attacks, Botnet, Infiltration, Heartbleed, etc.
- **Features:** 80+ flow-based attributes
    - Examples: Flow ID, Source IP, Destination IP, time-based stats, packet/byte counts
- **Visualization:** Attack distribution chart

Observations

- NSL-KDD: Balanced compared to original KDD, but still contains skewed classes.
- CICIDS2017: Strong imbalance (e.g., DDoS dominates).

Preprocessing Notes

- Encode categorical features (NSL-KDD).
- Normalize numerical values (CICIDS2017).
- Apply balancing methods (SMOTE).
- Perform feature selection for efficiency.