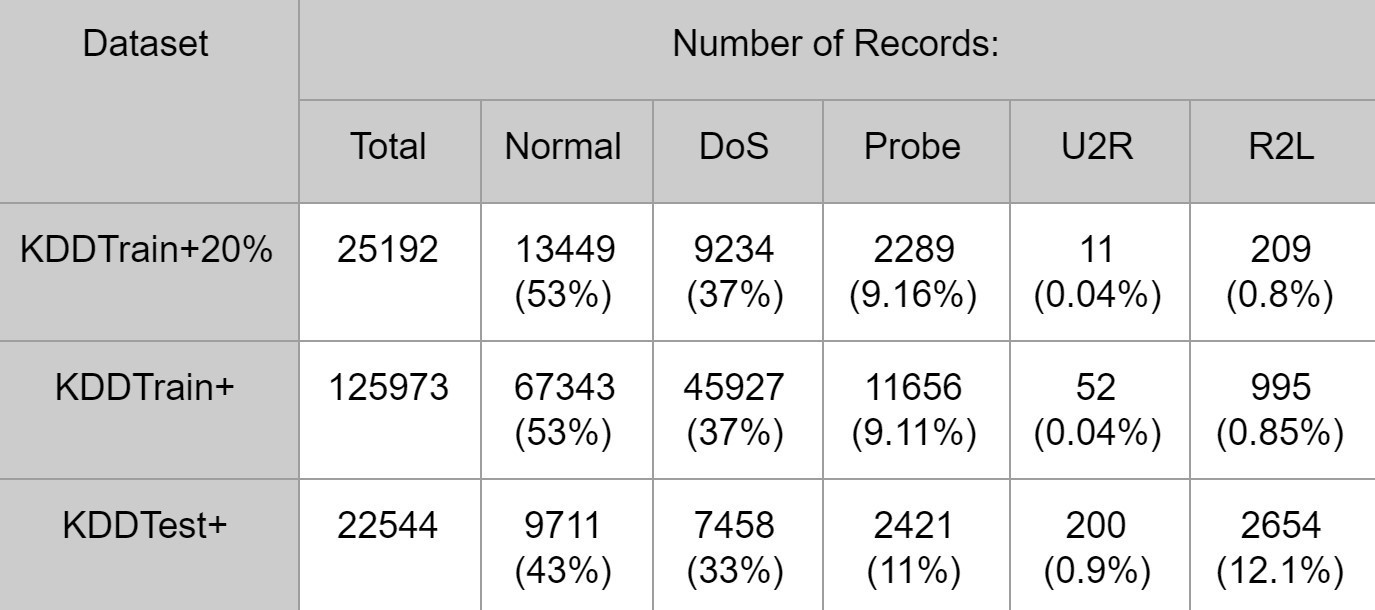
* No column names, so we have to hardcode all the column names by looking at the documentation
* Attack field names of attacks and “normal” for no attacks, so we replaced the values in the field with numerical values: 0 for a normal scenario and 1 for an attack scenario (for data analysis)
* NSL-KDD Dataset (for benchmarking):
  + DoS- Denial of service
  + Probing- Surveillance and other probing attacks
  + U2R- Unauthorized access to local superuser
  + R2L- Unauthorized access from a remote machine



* Attack vs Protocol:
  + Most attacks are targeting a specific protocol
  + satan, nmap and ipsweep are cross-protocol attacks
  + ICMP data is least frequently used in normal traffic
  + Pie charts show the distribution of attacks in a specific protocol
* Attack vs Services
  + A huge amount of normal traffic is on HTTP
  + Attacks are happening on other protocols which means attacks are searching for many different paths into systems
* Feature selection: Combination of protocol, flag and services tells a lot about the nature of the traffic. Then, coupling the duration of the connection will also profit us.
* Feature Engineering: We will observe and find features for multi-class classification