 **Networking/Security**: Focuses on detecting network intrusions (e.g., DoS, probing attacks), a core concern for network engineers and security experts.

 **Sitecore/Active Directory**: Intrusion detection applies to securing web platforms (Sitecore) and authentication systems (AD) by identifying malicious traffic (e.g., brute-force login attempts).

**Data Dictionary (NSL-KDD Dataset)**

The dataset has 41 features + 1 target (label). Below is a simplified summary of key features (full details on Kaggle):

| **Column Name** | **Data Type** | **Description** | **Possible Values/Range** |
| --- | --- | --- | --- |
| duration | integer | Length of connection (seconds). Longer durations may indicate attacks. | 0 to 58,329 |
| protocol\_type | string | Network protocol. Different protocols have varying attack risks. | 'tcp', 'udp', 'icmp' |
| service | string | Destination service (e.g., http, ftp). Some services are attack targets. | 70 categories (e.g., 'http', 'telnet') |
| flag | string | Connection status. Abnormal flags signal attacks. | 11 categories (e.g., 'SF', 'REJ') |
| src\_bytes | integer | Bytes sent from source. High values may indicate data exfiltration. | 0 to 1.3B |
| dst\_bytes | integer | Bytes sent to destination. Similar to src\_bytes. | 0 to 1.3B |
| logged\_in | binary | 1 if successfully logged in, 0 otherwise. Strong indicator of normal traffic. | 0 or 1 |
| count | integer | Connections to same host in last 2 seconds. High counts suggest attacks. | 0 to 511 |
| srv\_count | integer | Connections to same service in last 2 seconds. Similar to count. | 0 to 511 |
| same\_srv\_rate | float | % of connections to same service. Low rates may indicate probing attacks. | 0 to 1.0 |
| label | string | **Target**: Connection type (binary: 'normal' vs. 'anomaly'; or multiclass: 'normal' vs. specific attacks like 'smurf', 'neptune'). | Binary: 'normal', 'anomaly'; Multiclass: 23 categories |

* **Rows**: Train (~125,973), Test (~22,544).
* **Features**: 41 (3 categorical, 38 numerical/binary).
* **Target**: label (binary or multiclass classification).
* **Notes**: No missing values, but categorical variables (protocol\_type, service, flag) need one-hot encoding. Some features are highly correlated with the target (e.g., logged\_in, count), ensuring good model performance.

**Classification Tasks**

1. **Binary Classification**: Predict if a connection is 'normal' or 'anomaly' (attack). Practical for detecting malicious traffic in AD/Sitecore systems.
2. **Multiclass Classification**: Classify specific attack types (e.g., 'smurf' for DoS, 'satan' for probing). Useful for detailed security analysis.

**Usecase**:

* **Networking/Security**: Identify intrusions to protect network infrastructure.
* **Sitecore/AD**: Flag unauthorized access attempts (e.g., brute-force logins) or anomalous traffic to secure web apps or authentication systems.
* **Real-World**: Deploy models to firewalls or intrusion detection systems (IDS) for real-time monitoring.