

Documentação Técnica

Projeto: Análise de Atividades e Comportamentos de Ataques Cibernéticos

1. Visão Geral do Projeto

Este projeto tem como objetivo analisar padrões, comportamentos e características de ataques cibernéticos a partir do dataset **UNSW_NB15**, amplamente utilizado em estudos de segurança da informação. As análises buscam apoiar a identificação de vetores de ataque, serviços mais explorados, protocolos predominantes e métricas temporais relevantes, subsidiando decisões técnicas e estratégicas em cibersegurança.

As queries documentadas neste artefato cumprem três papéis fundamentais: - **Validação da qualidade dos dados** (detecção de valores nulos ou inconsistentes); - **Exploração descritiva** das variáveis críticas (tipo de ataque, serviço, protocolo); - **Geração de métricas analíticas** para entendimento do comportamento dos ataques.

2. Fonte de Dados

2.1 Tabela Principal: UNSW_NB15_TRAINING

Tabela contendo registros de tráfego de rede rotulados, com informações sobre conexões, protocolos, serviços e categorias de ataques.

Principais colunas utilizadas: - **ATTACK_CAT**: Categoria do ataque identificado; - **SERVICE**: Serviço de rede associado à conexão; - **PROTO**: Protocolo de comunicação utilizado; - **DUR**: Duração da conexão; - **CD**: Identificador único do registro.

2.2 Tabela Auxiliar: NEW

Tabela criada para corrigir tipagem inadequada da coluna DUR, originalmente importada como VARCHAR.

Colunas relevantes: - **CD**: Identificador único do registro (chave de junção); - **DUR**: Duração da conexão em formato numérico.

DocuSigned by:

Fábio Salve Menezes Júnior

3F5CD7525076464...

15/01/2026

3. Documentação das Queries

Query 1 – Análise da Categoria de Ataque (ATTACK_CAT)

3.1 Verificação de valores nulos

Objetivo: Avaliar a integridade dos dados da coluna ATTACK_CAT, identificando registros sem classificação de ataque.

Importância Analítica: Valores nulos podem comprometer análises estatísticas e modelos de detecção, além de indicar falhas no processo de rotulagem dos dados.

3.2 Listagem das categorias de ataque

Objetivo: Identificar todas as categorias distintas de ataques presentes no dataset.

Importância Analítica: Permite compreender o escopo dos ataques analisados e validar se o conjunto de dados contempla todos os tipos esperados (ex.: DoS, Exploits, Shellcode, Backdoor).

3.3 Quantificação de ataques do tipo Exploits

Objetivo: Calcular a frequência de ataques classificados como *Exploits*.

Importância Analítica: A categoria *Exploits* costuma representar ataques que exploram vulnerabilidades conhecidas. Quantificar sua ocorrência ajuda a medir o risco associado a falhas de software e sistemas desatualizados.

Query 2 – Análise por Serviço de Rede (SERVICE)

3.4 Verificação de valores nulos na coluna SERVICE

Objetivo: Identificar registros sem informação do serviço associado à conexão.

Importância Analítica: A ausência dessa informação prejudica análises que correlacionam ataques a serviços específicos, como HTTP, FTP ou DNS.

3.5 Listagem dos serviços existentes

Objetivo: Mapear todos os serviços distintos presentes no tráfego analisado.

Importância Analítica: Permite identificar superfícies de ataque mais expostas e serviços potencialmente críticos.

DocuSigned by:

Fábio Salve Menezes Júnior

3F5CD7525076464...

15/01/2026

3.6 Ataques DoS associados ao serviço HTTP

Objetivo: Quantificar ocorrências de ataques DoS direcionados especificamente ao serviço HTTP.

Importância Analítica: Ataques DoS contra HTTP impactam diretamente aplicações web. Essa métrica auxilia na priorização de mecanismos de mitigação, como rate limiting e WAFs.

Query 3 – Análise de Protocolos (PROTO)

3.7 Verificação de valores nulos na coluna PROTO

Objetivo: Detectar registros sem definição de protocolo de rede.

Importância Analítica: Protocolos são essenciais para entender o vetor técnico do ataque. Dados incompletos reduzem a confiabilidade da análise.

3.8 Listagem dos protocolos utilizados

Objetivo: Identificar todos os protocolos distintos observados no dataset.

Importância Analítica: Auxilia na identificação de padrões de ataque associados a protocolos específicos, como TCP ou UDP.

3.9 Protocolo mais comum em ataques Shellcode

Objetivo: Identificar o protocolo mais frequentemente utilizado em ataques do tipo *Shellcode*.

Importância Analítica: Ataques Shellcode costumam explorar execuções remotas. Conhecer o protocolo predominante apoia estratégias de detecção e monitoramento direcionado.

Query 4 – Análise de Duração dos Ataques (DUR)

3.10 Verificação de valores nulos na duração

Objetivo: Identificar registros sem informação de duração da conexão.

Importância Analítica: A métrica de duração é fundamental para análises comportamentais, como persistência e impacto do ataque.

3.11 Tipo de ataque com maior média de duração

Objetivo: Determinar qual categoria de ataque apresenta a maior duração média.

Aspectos Técnicos: - Uso de JOIN entre UNSW_NB15_TRAINING e NEW para acesso à duração numérica; - Cálculo de média com AVG; - Ordenação decrescente e retorno do maior valor.

Importância Analítica: Ataques mais longos tendem a ser mais persistentes e potencialmente mais danosos, exigindo atenção prioritária.

Query 5 – Análise Específica: Backdoor via FTP

3.12 Média de duração de ataques Backdoor utilizando FTP

Objetivo: Calcular a duração média de conexões associadas a ataques do tipo *Backdoor* quando o serviço utilizado é FTP.

Aspectos Técnicos: - Filtro por ATTACK_CAT = 'Backdoor' e SERVICE = 'ftp'; - Uso da tabela NEW para cálculo correto da média; - Agrupamento por tipo de ataque.

Importância Analítica: Ataques Backdoor via FTP podem indicar tentativas de acesso persistente ou exfiltração de dados. A análise de duração ajuda a entender o grau de comprometimento e o tempo de permanência do atacante.

4. Considerações Finais

As queries documentadas constituem a base exploratória e analítica do projeto de Análise de Atividades e Comportamentos de Ataques Cibernéticos. Elas garantem: - Confiabilidade dos dados analisados; - Visibilidade sobre padrões técnicos de ataque; - Subsídios objetivos para tomada de decisão em segurança da informação.

Este conjunto pode ser evoluído para camadas mais avançadas, como modelagem preditiva, correlação temporal e integração com dashboards analíticos.

DocuSigned by:

Fábio Salve Menezes Júnior

3F5CD7525076464...

15/01/2026