NETWORK INTRUSION DETECTION

MINERAÇÃO DE DADOS

Grupo 6:

- Ana Murta (pg50184)
- Hugo Gomes (pg51242)
- Manuel Novais (pg50575)

MOTIVAÇÃO

- Aumento do número de ataques cibernéticos
- Dependência crescente da tecnologia
- Crescente importância da privacidade dos dados

OBJETIVOS

Criar um sistema de deteção de *network intrusion* que deverá conseguir categorizar cada conexão dum conjunto destas como normal ou anomalia.

Futuro: o sistema ser usado em tempo real para proteção de ambiente

- Identificar padrões de tráfego de rede maliciosos
- Identificar anomalias no tráfego de rede
- Identificar comportamentos de usuários maliciosos e fontes de ameaças

FONTE DE DADOS

- https://www.kaggle.com/datasets/sampadab17/network-intrusion-detection.

- https://www.kaggle.com/datasets/h2020simargl/simargl2021-network-intrusion-detection-dataset.

kaggle

DESCRIÇÃO DOS DADOS

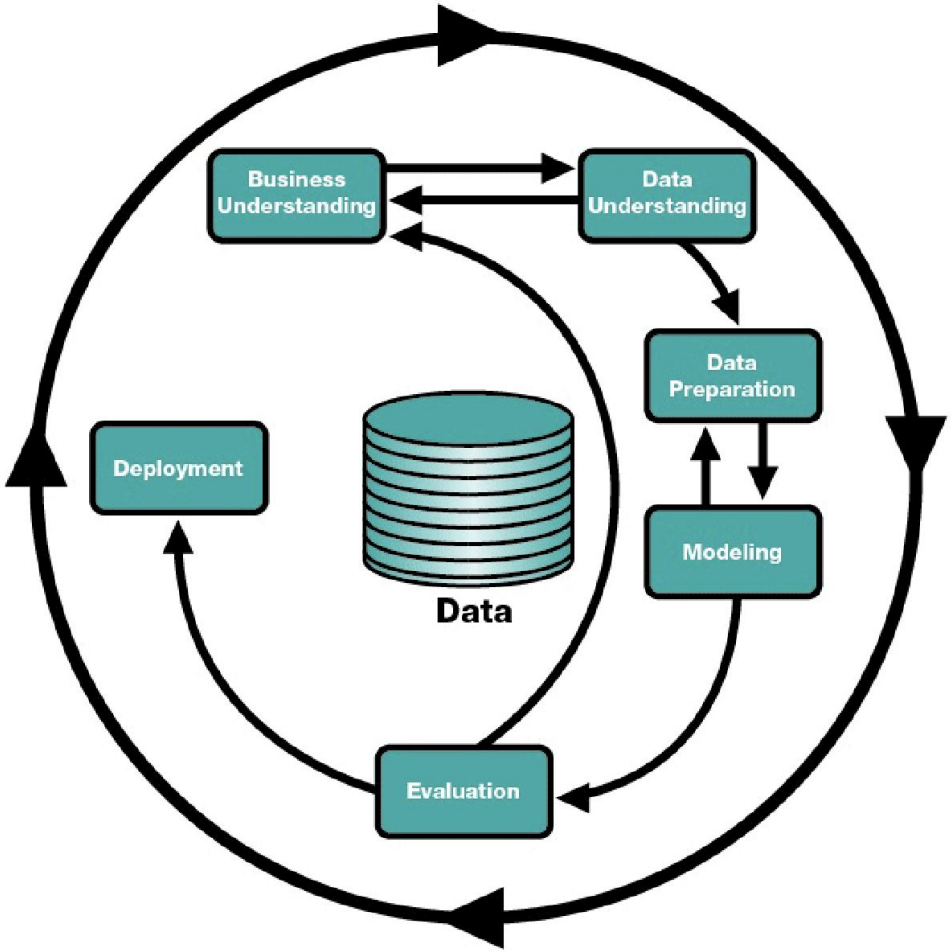
Network Intrusion Detection

- Variedade de intrusões simuladas num ambiente de rede militar.
- Dados de pacotes TCP/IP de uma rede simulando uma LAN da Força aérea dos USA.
- Oferece um ambiente realista com o objetivo de ser bombardeada com múltiplos ataques.
- 42 colunas (incluindo a label).

SIMARGL2021

- Tráfego real recolhido após terem sido feitos ataques a uma rede real.
- 50 colunas (incluindo a label)
- Dividido em 4 partes.

METODOLOGIA



- CRIPS-DM

TAREFAS A DESENVOLVER

1 **Recolha dos dados**

Os datasets vão ser importados

2 **Exploração dos dados**

Os dados vão ser explorados através de extração de dados estatísticos e construção de gráficos.

3 **Tratamento dos dados**

Os dados vão ser tratados e vão ser aplicadas técnicas de *feature engineering* e transformação de dados para obter as *features* mais importantes a ser usadas para treinar os modelos.

4 **Treino e validação dos modelos de aprendizagem**

Vão ser treinados e testados vários modelos de aprendizagem de forma a encontrar o modelo mais adequado ao problema.

FERRAMENTAS

Pandas

Importe e tratamento dos dados

Jupyter Notebook

Implementação do código

Sklearn

Implementação de algoritmos de *machine learning*

Seaborn

Visualização dos dados

ALGORITMOS

- Decision Tree Classifier
- Naive Bayes Model
- K Nearest Neighbors (KNN) classification model
- Random Forest Classifier
- Logistic Regression Model

NETWORK INTRUSION DETECTION

MINERAÇÃO DE DADOS

Grupo 6:

- Ana Murta (pg50184)
- Hugo Gomes (pg51242)
- Manuel Novais (pg50575)