Estudo da Influência das Features e do Desempenho dos Classificadores na Detecção de Anomalias em Redes

João Lucas Oliveira Mota[1], João Lucas Rodrigues da Silva[2], José Alberto Rodrigues Neto[2]

[1] Departamento de Engenharia de Teleinformática – Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brasil

[2]Departamento de Computação – Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brasil

{jlucasoliveira2002}@gmail.com, {joaolucas.rodrigues,jose.alberto}@alu.ufc.br

Abstract. *This document is a model and instructions for LaTeX. This and the sbc-template style define the components of your paper. Do not use symbols, special characters, footnotes, or math in the paper title or abstract.*

Resumo. *Este documento apresenta o modelo e o conjunto de instruções para artigos elaborados em LaTeX segundo o padrão da SBC. O arquivo sbc-template define os principais componentes do artigo, como título, autores e seções.*

1. Introdução

CHANGE ME Este artigo investiga a influência da seleção de atributos e do desempenho de diferentes classificadores na detecção de anomalias em tráfego de redes de computadores.

2. Dataset e pré-processamento

2.1. Dataset

O conjunto de dados utilizado foi obtido a partir da plataforma Kaggle, no repositório *Network Intrusion Detection*, disponibilizado publicamente por Sampada Bhosale [Bhosale 2018].

O dataset foi construído a partir da simulação de um ambiente de rede militar típico da Força Aérea dos Estados Unidos, desenvolvido para representar diferentes tipos de tráfego de rede, incluindo tanto comunicações legítimas quanto diversos tipos de ataques. Cada instância do conjunto de dados representa uma conexão de rede, contendo atributos como a quantidade de bytes transmitidos pela origem, a duração da conexão e outros parâmetros, conforme descrito na tabela apresentada no Apêndice.

Por fim, após as simulações realizadas de forma supervisionada, as conexões foram classificadas como tráfego normal ou como intrusão, fazendo com que a variável alvo seja do tipo binária, assumindo os valores Normal ou Anomalia.

2.2. Pré-processamento

CHANGE ME A aplicação de PCA como etapa de pré-processamento tem se mostrado eficaz para reduzir a dimensionalidade dos dados sem perda significativa de informação, preservando cerca de 99% da variância original mesmo com reduções superiores a 50% no número de atributos, conforme observado por Santos e Miani [Santos and Miani 2025].

Agradecimentos

Os autores agradecem à Universidade Federal do Ceará (UFC) pelo suporte institucional.

References

Bhosale, S. (2018). Network intrusion detection. Kaggle. Acesso em 2025.

Santos, K. C. and Miani, R. S. (2025). Impacto da redução de dimensão e seleção de atributos na generalização de modelos de detecção de intrusão. In *Anais do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, Uberlândia, MG, Brazil. SBC.

Apêndice A — Descrição Completa do Dataset

Table 1. Descrição completa dos atributos do dataset de intrusão em redes

Atributo	Descrição
duration	Duração da conexão de rede
protocol_type	Protocolo utilizado na conexão (TCP, UDP ou ICMP)
service	Serviço de rede acessado
flag	Estado da conexão conforme o protocolo
src_bytes	Quantidade de bytes enviados pela origem
dst_bytes	Quantidade de bytes enviados pelo destino
land	Indica se origem e destino possuem mesmo IP e porta
wrong_fragment	Número de fragmentos incorretos
urgent	Número de pacotes marcados como urgentes
hot	Indicadores de comportamentos suspeitos
num_failed_logins	Número de tentativas de login malsucedidas
logged_in	Indica se o login foi realizado com sucesso
num_compromised	Número de condições comprometidas
root_shell	Indica obtenção de shell com privilégio root
su_attempted	Tentativas de uso do comando *su*
num_root	Número de acessos root
num_file_creations	Número de arquivos criados
num_shells	Número de shells abertos
num_access_files	Número de acessos a arquivos sensíveis
num_outbound_cmds	Número de comandos externos enviados
is_host_login	Indica login como host
is_guest_login	Indica login como convidado
count	Conexões com o mesmo host em janela de tempo
srv_count	Conexões com o mesmo serviço em janela de tempo
serror_rate	Taxa de erros SYN
srv_serror_rate	Taxa de erros SYN para o serviço
rerror_rate	Taxa de erros de resposta
srv_rerror_rate	Taxa de erros de resposta para o serviço
same_srv_rate	Taxa de conexões para o mesmo serviço
diff_srv_rate	Taxa de conexões para serviços diferentes
srv_diff_host_rate	Taxa de serviços acessando hosts distintos
dst_host_count	Número de conexões para o mesmo host destino
dst_host_srv_count	Número de conexões para o mesmo serviço no host
dst_host_same_srv_rate	Taxa de serviços iguais para o host destino
dst_host_diff_srv_rate	Taxa de serviços diferentes para o host destino
dst_host_same_src_port_rate	Taxa de conexões com mesma porta de origem
dst_host_srv_diff_host_rate	Taxa de serviços com hosts distintos
dst_host_serror_rate	Taxa de erros SYN no host destino
dst_host_srv_serror_rate	Taxa de erros SYN por serviço no host destino
dst_host_rerror_rate	Taxa de erros de resposta no host destino
dst_host_srv_rerror_rate	Taxa de erros de resposta por serviço no host destino