

Relatório de remoção de linhas de Do fluxos IOT e Normal

Neste relatório visa entender as features que estão sendo removidas do fluxo gerado via nfstream, neste relatório separaremos o relatório em duas partes IOT e Normal para sabermos se temos o mesmo comportamento de número de features com valores nulos para o tráfego IOT e o tráfego normal.

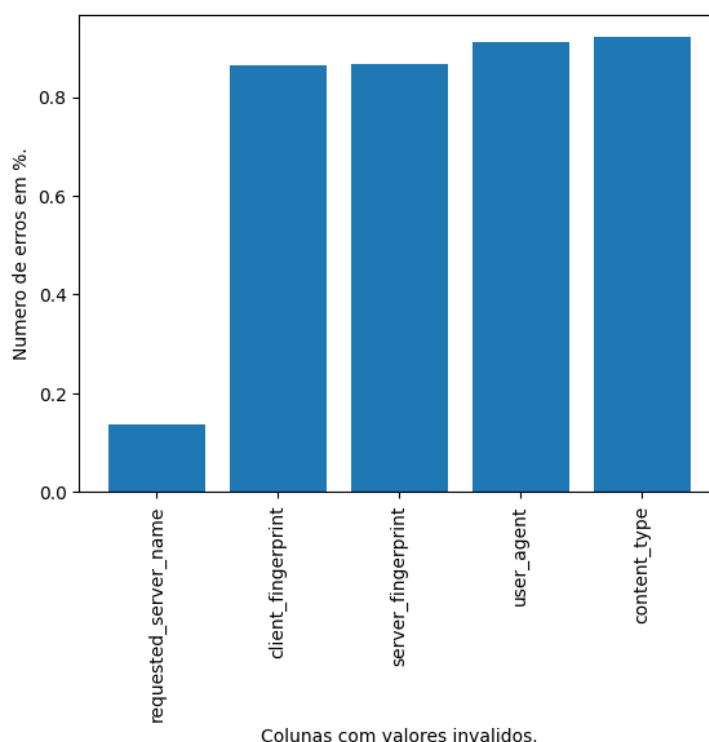
Em ambos o tratamento de erros foi feito em duas partes na primeira removemos as features que possuem 'application', 'ip', 'mac', que são as features:

- application_is_guessed;
- application_name;
- dst_ip;
- dst_mac;
- src_ip;
- src_mac.

No tráfego normal e IOT, algumas colunas apresentaram um número muito elevado de valores nulos e que na hora de excluir as linhas com valores nulos estava gerando problema já que o data frame inteiro estava sendo removido, o que me levou a pegar essas colunas com mais de 50% de valores inválidos e removê-las para ai sim deletar linhas com valores nulos, as colunas em questão são:

- client_fingerprint;
- content_type;
- requested_server_name;
- server_fingerprint;
- user_agent.

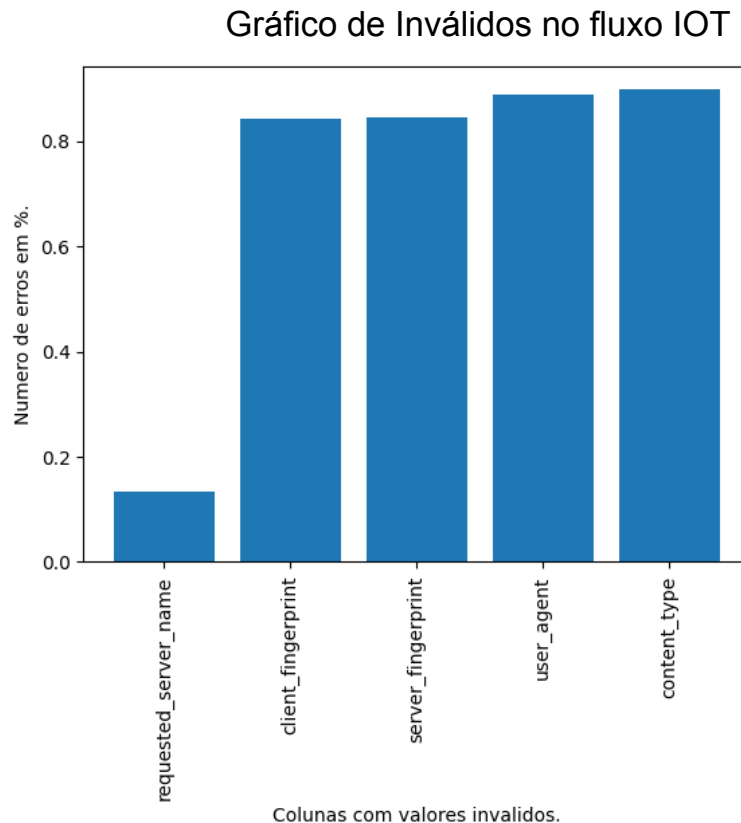
Gráfico de Inválidos no fluxo normal



Depois foi removido as linhas que possuem valores inválidos que foram 55644 de um total de 415021 linhas, que pode ser melhor visualizado no relatório de limpeza do tráfego normal.

Tráfego IOT

No Tráfego IOT tivemos o mesmo comportamento descrito acima com as mesmas colunas tiveram um índice elevado de valores inválidos como indica o gráfico abaixo:



O relatório do fluxo IOT Possui o mesmo problema de número muito elevado de erros nas mesmas colunas descritas acima no tráfego Normal, já na limpeza dos nulos foram removidas 29634 de um total de 404571 linhas que pode melhor ser visualizado no relatório de limpeza do IOT.

Conclusão

Parece que nas colunas que estão gerando erro descritas acima, os erros são causados por uma particularidade no fluxo em que não é possível determinar um valor para a coluna a depender do fluxo já que a porcentagem de erros para cada coluna e a mesma do tráfego IOT e normal, no relatório junto com o html dos fluxos IOT e o Normal antes e depois da limpeza, será enviando também das linhas excluídas durante a limpeza de valores nulos sem levarem conta as linha excluídas juntamente com o csv que origem de cada html do relatório.