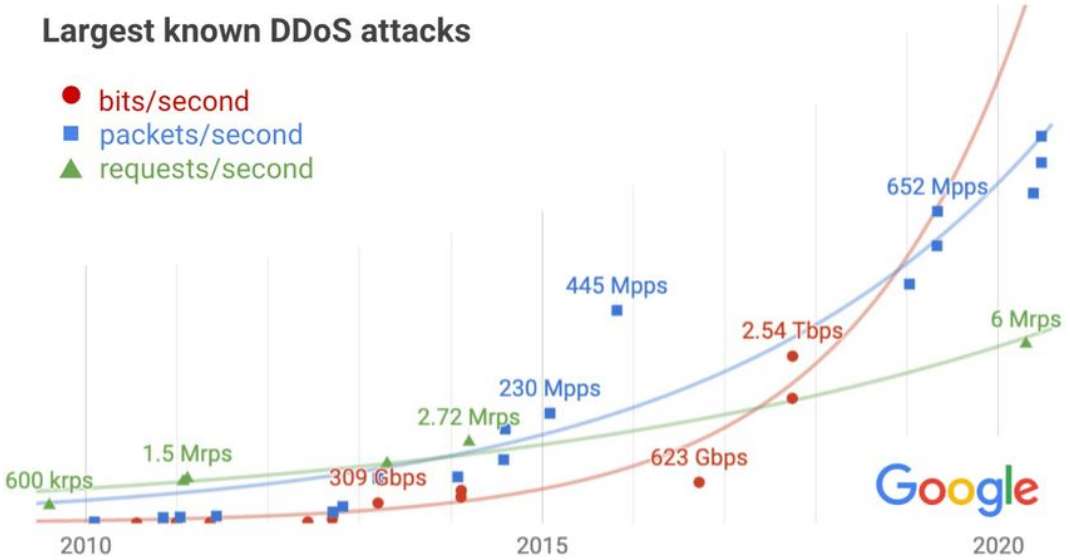
Uso De Aprendizado de Máquina Para Detecção De Ataque s DDoS

Orientador: Prof Dr. Kelton Augusto Pontara da Costa
Orientando: Gustavo Amaral Duarte Rego
RA: 201025817

Problemática

Ataques DDoS

- Ataques DDoS estão cada vez mais comuns e poderosos com o passar do tempo.
- Técnicas mais avançadas de detecção de ataques são necessárias conforme a escala aumenta.
- Redes neurais são utilizadas de forma recorrente para reconhecimento de padrões.



(CHAGANTI; BHUSHAN; RAVI, 2022)

Objetivos

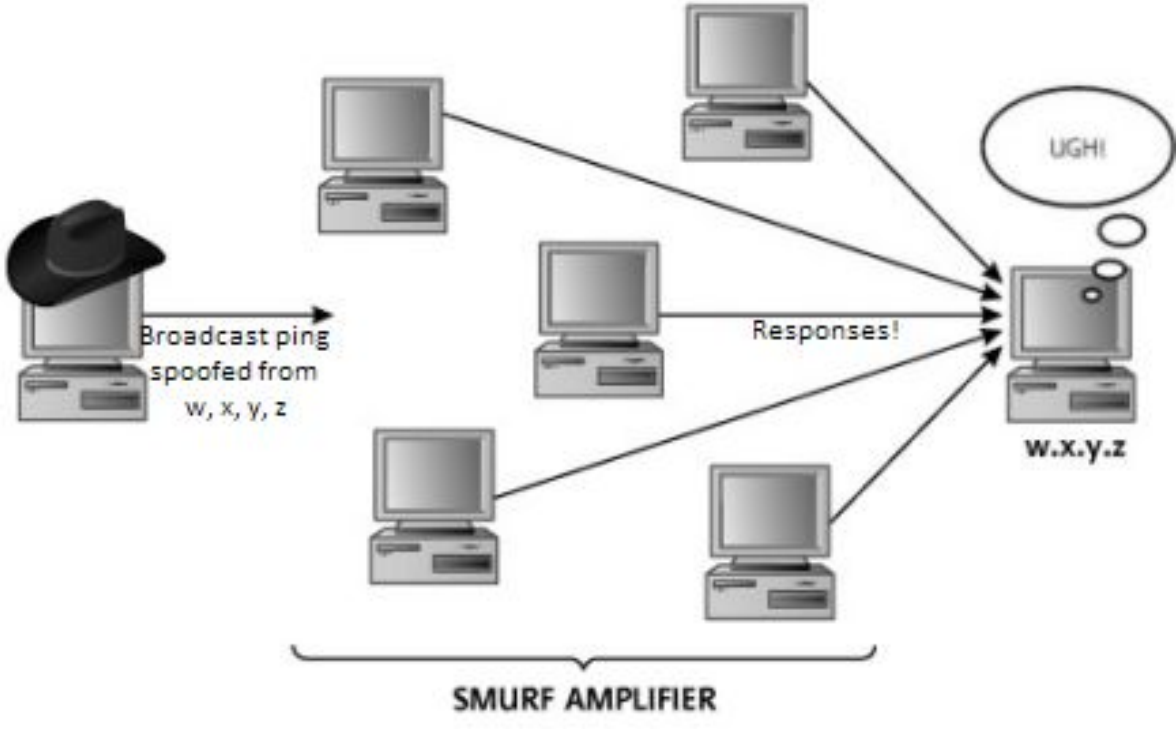
- Construir um algoritmo capaz de detectar ataques DDoS com base em técnicas de Aprendizado de máquina .
- Analisar os resultados obtidos e testar a eficiência da rede neural.

Fundamentação Teórica

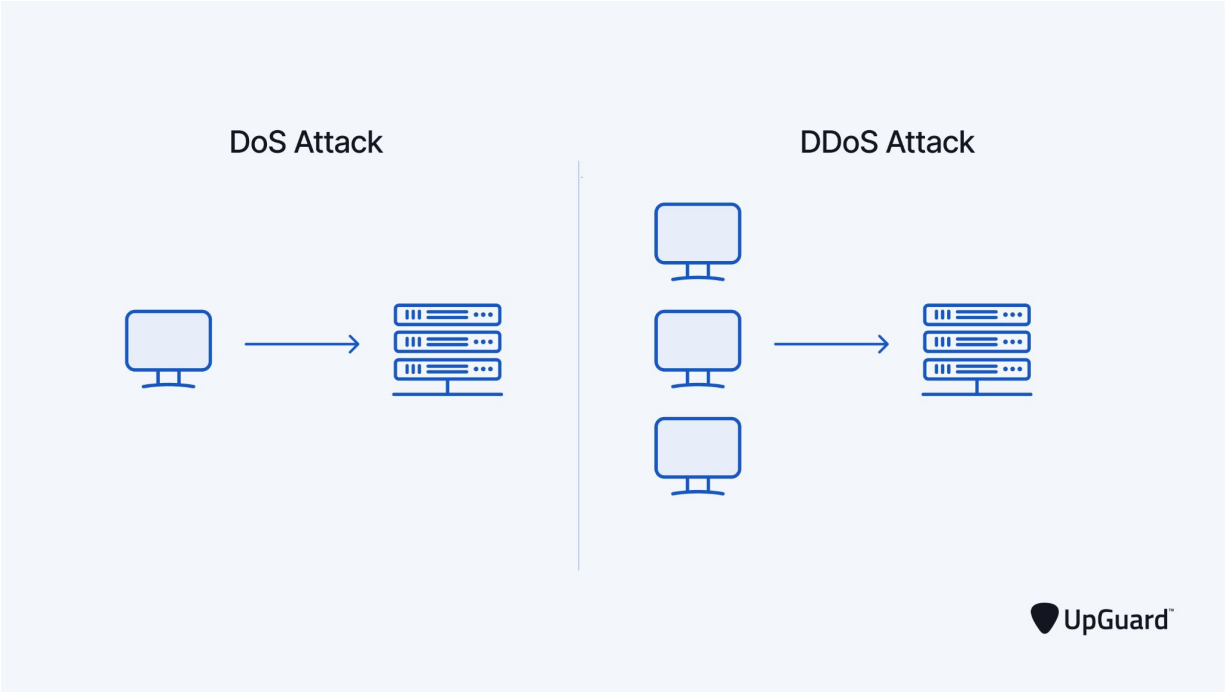
Ataques DDoS

- (HOQUE; BHATTACHARYYA; KALITA, 2015), Um ataque DDoS é um ataque coordenado que utiliza muitos *hosts* comprometidos.
- Objetivo de esgotar recursos de infraestrutura (Rede ou Hardware).
- É necessário um grande número de *hosts* para que o ataque seja eficiente.

Ataques DDoS



Ataques DDoS

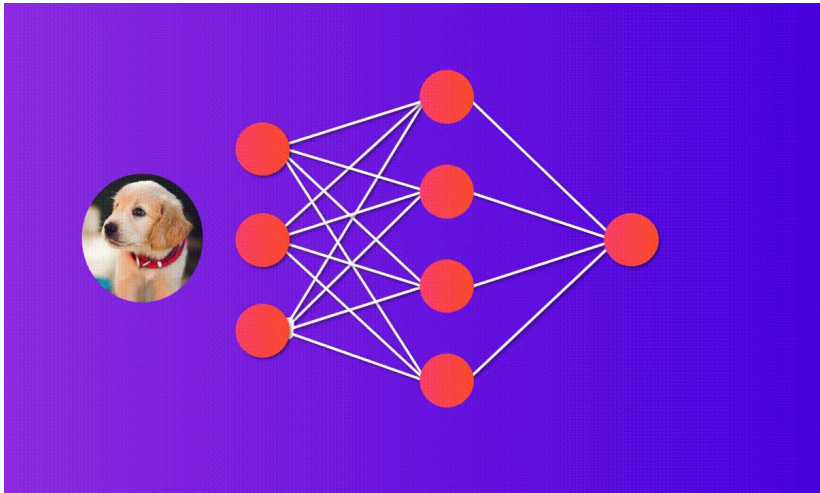


Aprendizado de Máquina

- Área da Inteligência Artificial voltada a criar algoritmos capazes de adquirir conhecimento.
- Tipos de algoritmos diferentes para aprendizado de máquina (*Support Vector Machines*, Redes Neurais Artificiais, entre outros).
- Redes Neurais Artificiais normalmente utilizam múltiplas camadas pois esses modelos conseguem discernir mais nuances entre a camada de entrada e saída.

Aprendizado de Máquina

- Redes Neurais Artificiais
  - Predizem um resultado baseado no treinamento.



Materiais e métodos

Materiais e métodos

- Hardware utilizado
  - Notebook pessoal
    - Intel Core i5 10300H
    - NVidia GTX 1650
    - 32GB RAM 3200MHz
    - 1TB SSD NVMe
  - Google Colab (free)

Materiais e métodos

- Software utilizado
    - Visual Studio Code - Jupyter

Materiais e métodos

- Foi selecionado o algoritmo de Multilayer Perceptron para a confecção do modelo.
- 1 camada de entrada com 64 neurônios, 1 camada intermediária com 32 neurônios e uma camada de saída para binarizar o resultado.

Materiais e métodos

- Conjunto de dados utilizado:
  - (DEVENDRA416, 2020), com 10 Milhões de amostras e 87 parâmetros.
  - Tráfego de rede rotulado com parâmetros que podem ser “DDoS” ou “benigno”.

Materiais e métodos

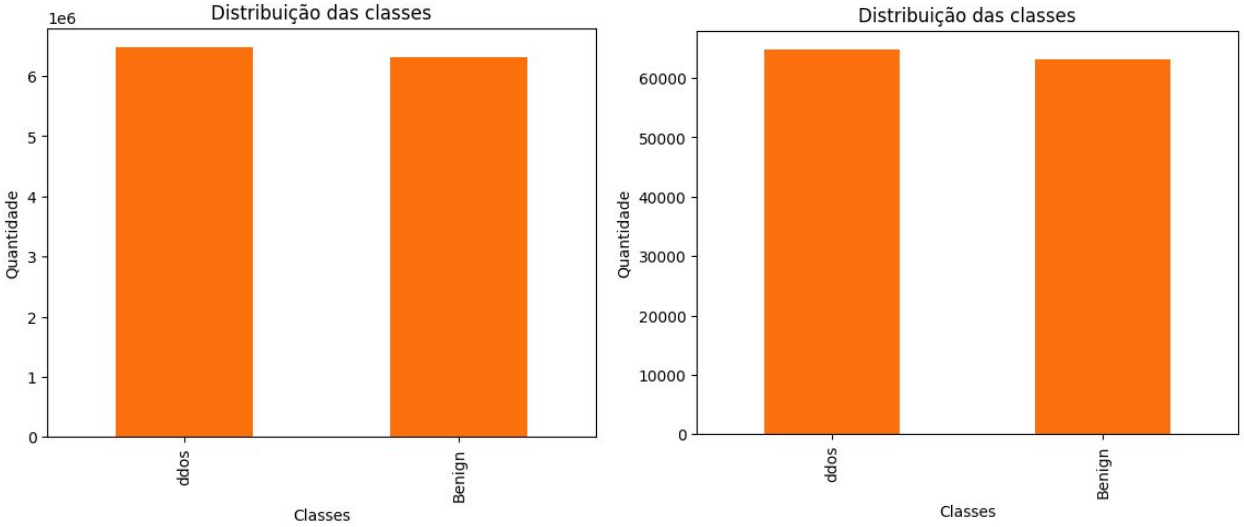
- Pré-processamento dos dados
    - O conjunto foi dividido em 5 subconjuntos diferentes, com 100 mil, 250 mil, 500 mil 1 Milhão e 10 Milhões de dados.

Materiais e métodos

- Pré-processamento dos dados
    - Por razões de armazenamento, foram escolhidos 4 parâmetros principais para o treinamento da rede.
        - IP de Destino
        - Tamanho médio dos pacotes
        - Pacotes Por Segundo

Materiais e métodos

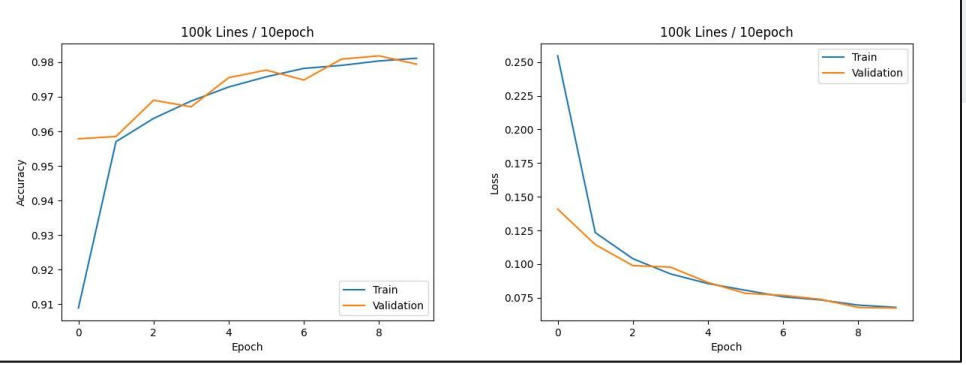
- Pré-processamento dos dados



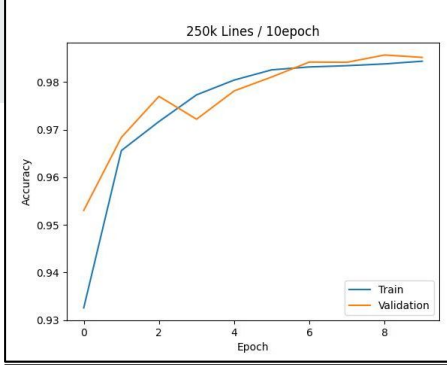
Análises e Resultados

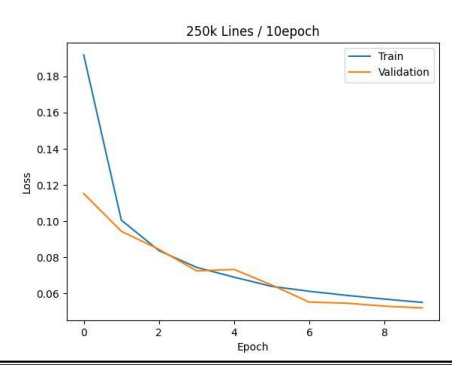
Análises e resultados

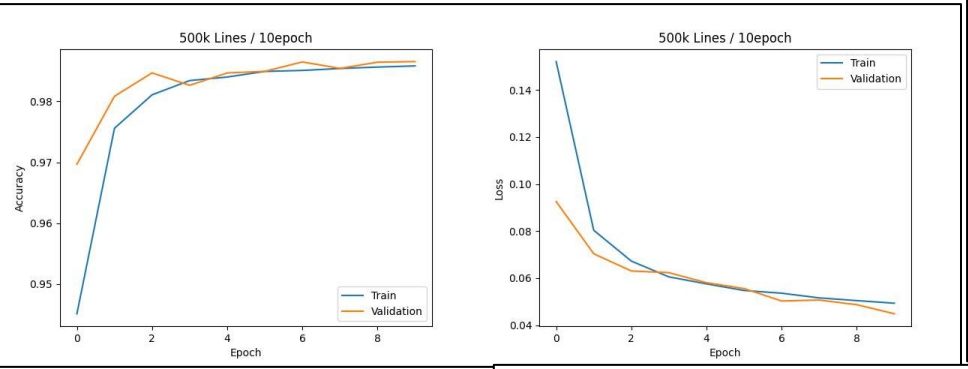
- Com os 5 conjuntos de dados foi possível obter as seguintes curvas de acurácia e perda:


100k Lines / 10epoch
Accuracy
Epoch
Train
Validation

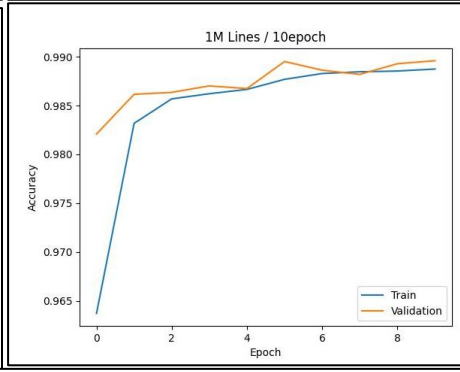
100k Lines / 10epoch
Loss
Epoch
Train
Validation

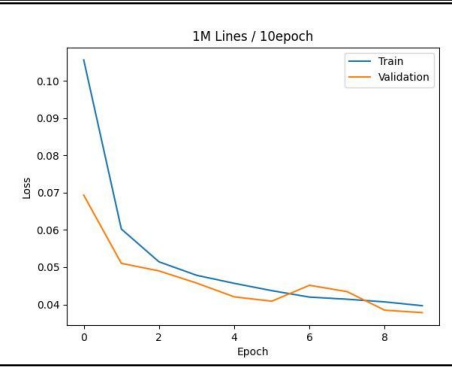

250k Lines / 10epoch
Accuracy
Epoch
Train
Validation

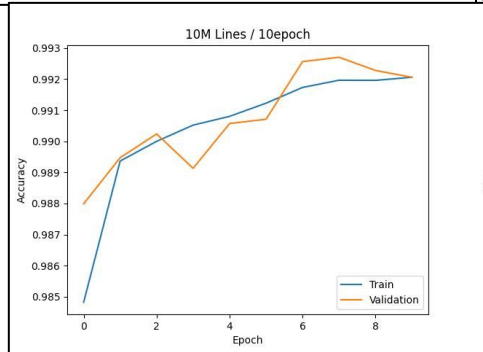

250k Lines / 10epoch
Loss
Epoch
Train
Validation

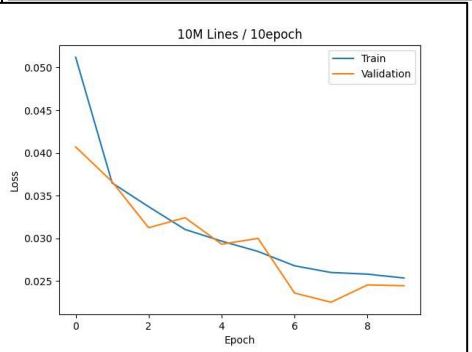

500k Lines / 10epoch
Accuracy
Epoch
Train
Validation

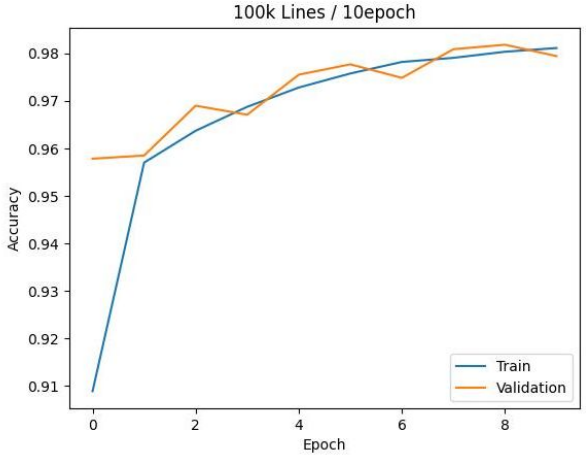
500k Lines / 10epoch
Loss
Epoch
Train
Validation

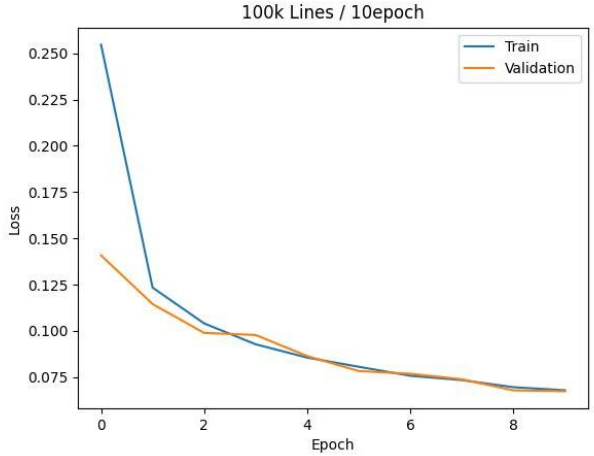

1M Lines / 10epoch
Accuracy
Epoch
Train
Validation


1M Lines / 10epoch
Loss
Epoch
Train
Validation

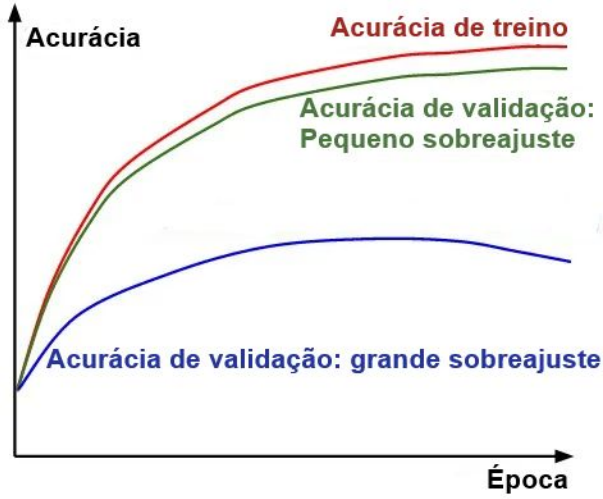

10M Lines / 10epoch
Accuracy
Epoch
Train
Validation


10M Lines / 10epoch
Loss
Epoch
Train
Validation


100k Lines / 10epoch
Accuracy
Epoch
Train
Validation

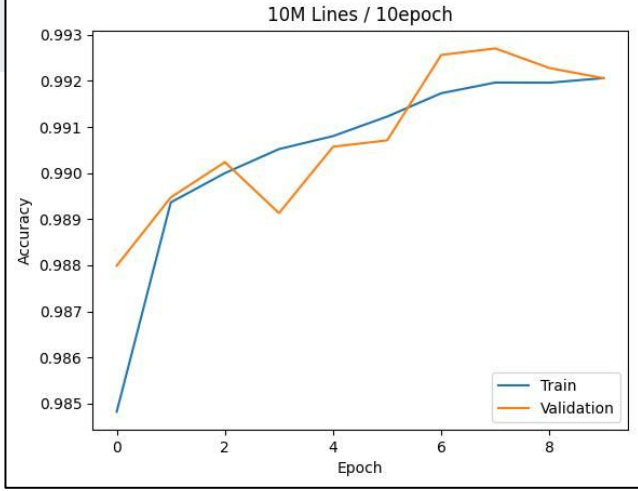

100k Lines / 10epoch
Loss
Epoch
Train
Validation

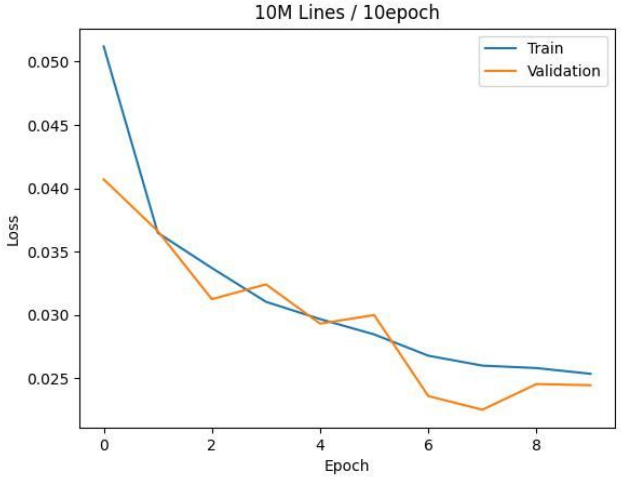

Perda
Taxa de aprendizado muito alta
Baixa taxa de aprendizado
Alta taxa de aprendizado
Boa taxa de aprendizado
Época

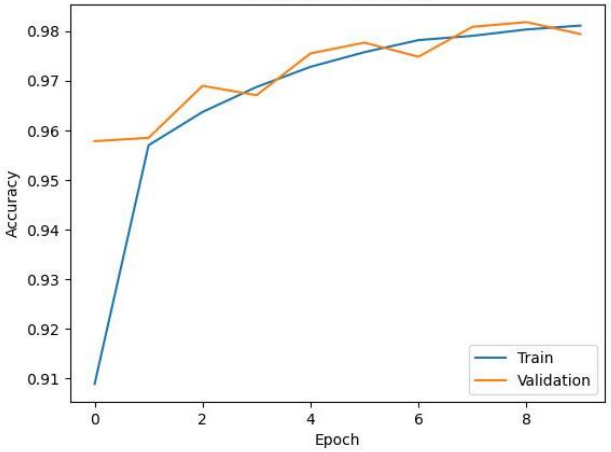

Acurácia
Acurácia de treino
Acurácia de validação:
Pequeno sobreajuste
Acurácia de validação: grande sobreajuste
Época

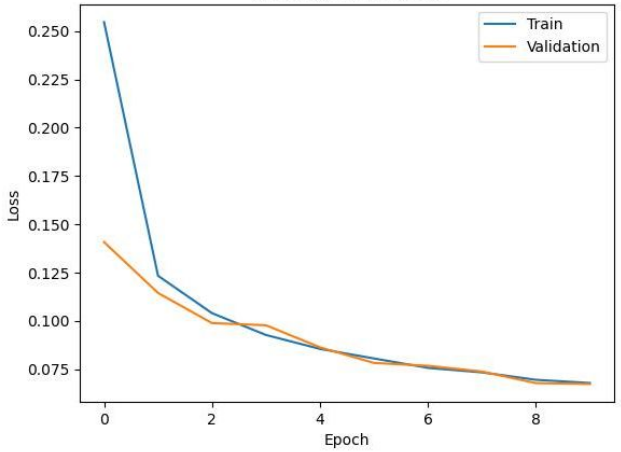
Análises e resultados

- Ao observar os gráficos anteriores, é possível inferir que a quantidade de dados impacta sim no resultado, porém, para o caso testado, não significativamente.
- Existe o ganho de 1% em termos de acurácia, porém para isso, foi necessário aumentar em 100 vezes a quantidade de dados.


10M Lines / 10epoch
Accuracy
Epoch
Train
Validation


10M Lines / 10epoch
Loss
Epoch
Train
Validation


100k Lines / 10epoch
Accuracy
Epoch
Train
Validation


100k Lines / 10epoch
Loss
Epoch
Train
Validation

Trabalhos futuros

- Utilização de outras técnicas de Aprendizado de Máquina para comparação;
- Modificar os parâmetros selecionados para avaliar como o modelo se ajusta;
- Expandir o escopo do modelo para reconhecimento de outros ataques cibernéticos;
- Testar o modelo em um ambiente real, com um tráfego do “ dia a dia”.

Referências

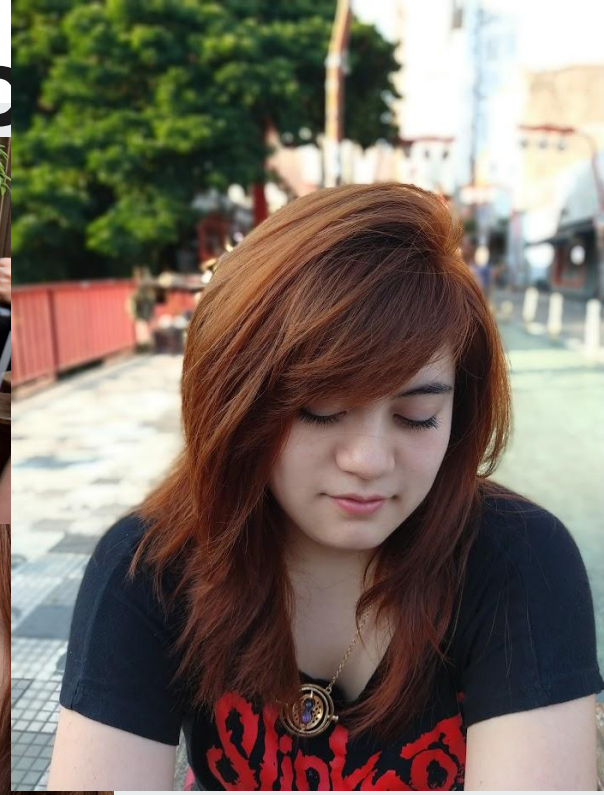
HOQUE, N.; BHATTACHARYYA, D. K.; KALITA, J. K. Botnet in ddos attacks: Trends and challenges. IEEE Communications Surveys Tutorials, v. 17, n. 4, p. 2242–2270, 2015. Disponível em: https://ieeexplore.ieee.org/abstract/document/7160662. Acesso em: 01 nov. 2023.

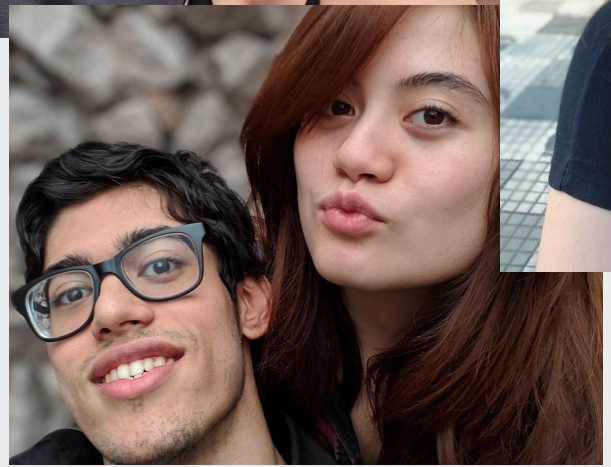
CHAGANTI, R.; BHUSHAN, B.; RAVI, V. The role of blockchain in ddos attacks mitigation: techniques, open challenges and future directions. 02 2022. Disponível em: https://www.researchgate.net/publication/358458534_The_role_of_Blockchain_in_DDoS_attacks_mitigation_techniques_open_challenges_and_future_directions. Acesso em: 02 set. 2023.

SHEN, B. W.; SETHI, G.; ZAKKA, K.; MOINDROTA, O.; DARDNER, R.; KULAL, S. CS231n Convolutional Neural Networks for Visual Recognition. 2020. Disponível em: https://cs231n.github.io/neural-networks-3/. Acesso em: 29 out. 2023.

Agradecimentos

Agradecimento





Agradecimentos